

AI and Media: Skills, Methods, and Applications

Prof. Stefan Feuerriegel

Institute of AI in Management
LMU Munich
<https://www.ai.bwl.lmu.de>



ABOUT OUR INSTITUTE

Solving real-world problems with artificial intelligence (AI)



What defines our research

1 Information

We solve management problems of **relevance** by using data science











2 Innovation

We develop **new** algorithms from the area of AI (statistics, computer science, etc.)

3 Impact

We evaluate the added value of our tools **rigorously** in management practice

Research Team

 <p>Dominik Bär, M.Sc. PhD Candidate and Research Assistant Send an email More</p>	 <p>Dennis Frauen, M.Sc. PhD Candidate and Research Assistant Send an email More</p>
 <p>Dominique Geißler, M.Sc. PhD Candidate and Research Assistant Send an email More</p>	 <p>Konstantin Heß, M.Sc. PhD Candidate and Research Assistant Send an email More</p>
 <p>Yuchen Ma, M.Sc. PhD Candidate and Research Assistant Send an email More</p>	 <p>Abdurahman Maarouf, M.Sc. PhD Candidate and Research Assistant Send an email More</p>
 <p>Valentin Melnychuk, M.Sc. PhD Candidate and Research Assistant Send an email More</p>	 <p>Simon Schallmoser, M.Sc. PhD Candidate and Research Assistant Send an email More</p>
 <p>Mareen Schröder, M.Sc. PhD Candidate and Research Assistant Send an email More</p>	 <p>Jonas Schweisthal, M.Sc. PhD Candidate and Research Assistant Send an email More</p>

Experimentation



Generative AI



Causal AI

If it bleeds,
it leads?



Negativity drives online news consumption



- $N = 22,743$ randomized controlled trials (each with around 4 headlines)



- Key variable: click-through rate $CTR_{ij} = \frac{\text{clicks}_{ij}}{\text{impressions}_{ij}}$

Sentiment analysis

Text

KRONES AG: KRONES' growth continues strong in the first three quarters of

Krones AG / Quarter Results

During the first nine months of 2008 KRONES remained on course for growth, despite the cyclical downturn. On a like-for-like basis, sales rose by 12.5 % to reach Euro 1,765.9 m. During the period under review, the company benefited from the increasing number of clients looking for all-inclusive job packages. Another growth driver during the year's first three quarters was the group's Plastic Technology Division. KRONES is the world's leading vendor of machines and [...]

Markings: examples of text components relevant for investors

Positive

Negative

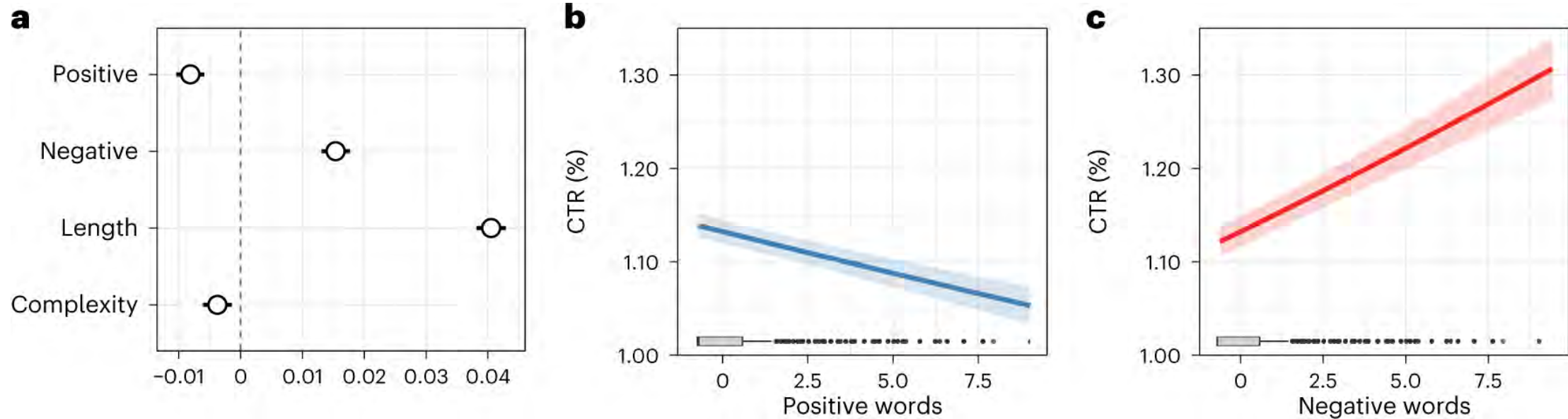
Extracting the positivity or negativity

- Frequencies of **dictionary** words
 - e.g. NRC emotion lexicon
- Resulting variable

$$\text{Positive}_{ij} = \frac{n_{\text{positive}}}{n_{\text{total}}} \text{ and } \text{Negative}_{ij} = \frac{n_{\text{negative}}}{n_{\text{total}}}$$

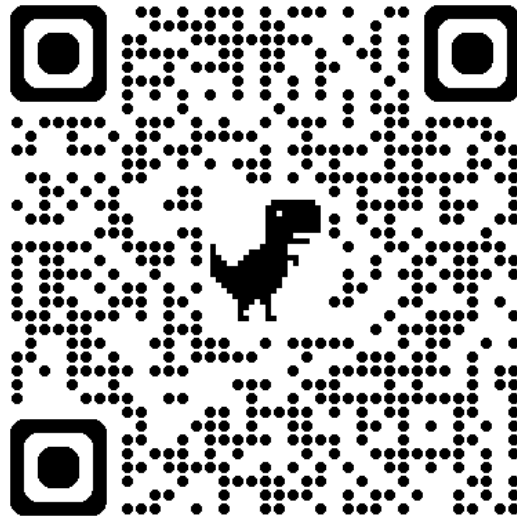
- Standardization

The effect of positive and negative words in news headlines on the CTR



Headlines ($N = 53,669$) were examined over 12,448 RCTs. **a**, Estimated standardized coefficients (circles) with 99% confidence intervals (error bars) for positive and negative words and for further controls. The variable 'PlatformAge' is included in the model during estimation but not shown for better readability. Full estimation results are in Supplementary Table 3. **b, c**, Predicted marginal effects on the CTR (lines). The error bands (shaded area) correspond to 99% confidence intervals. Boxplots show the distribution of the variables in our sample (centre line gives the median, box limits are upper and lower quartiles, whiskers denote minimum/maximum, points are outliers defined as being beyond 1.5× the interquartile range).

Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., & Feuerriegel, S. (2023). Negativity drives online news consumption. *Nature Human Behaviour*, 7(5), 812-822.



Experimentation



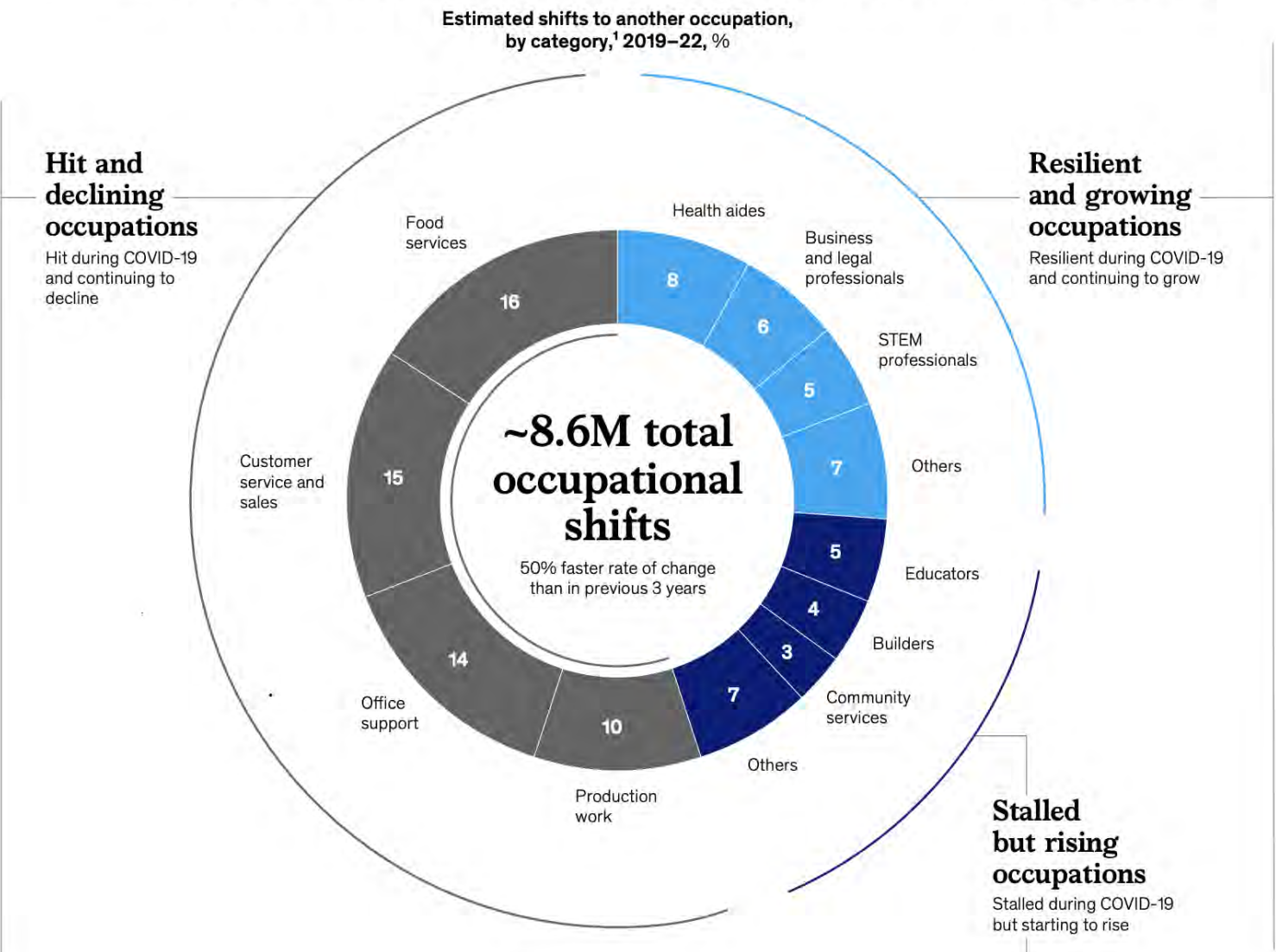
Generative AI



Causal AI

But what about journalists?

More than 50 percent of recent occupational shifts in the United States involved workers leaving roles in food services, customer service, office support, and production.



Source: McKinsey (2023)

Research questions



RQ 1: How does prompt engineering training influence the user experience of journalists when interacting with LLM?

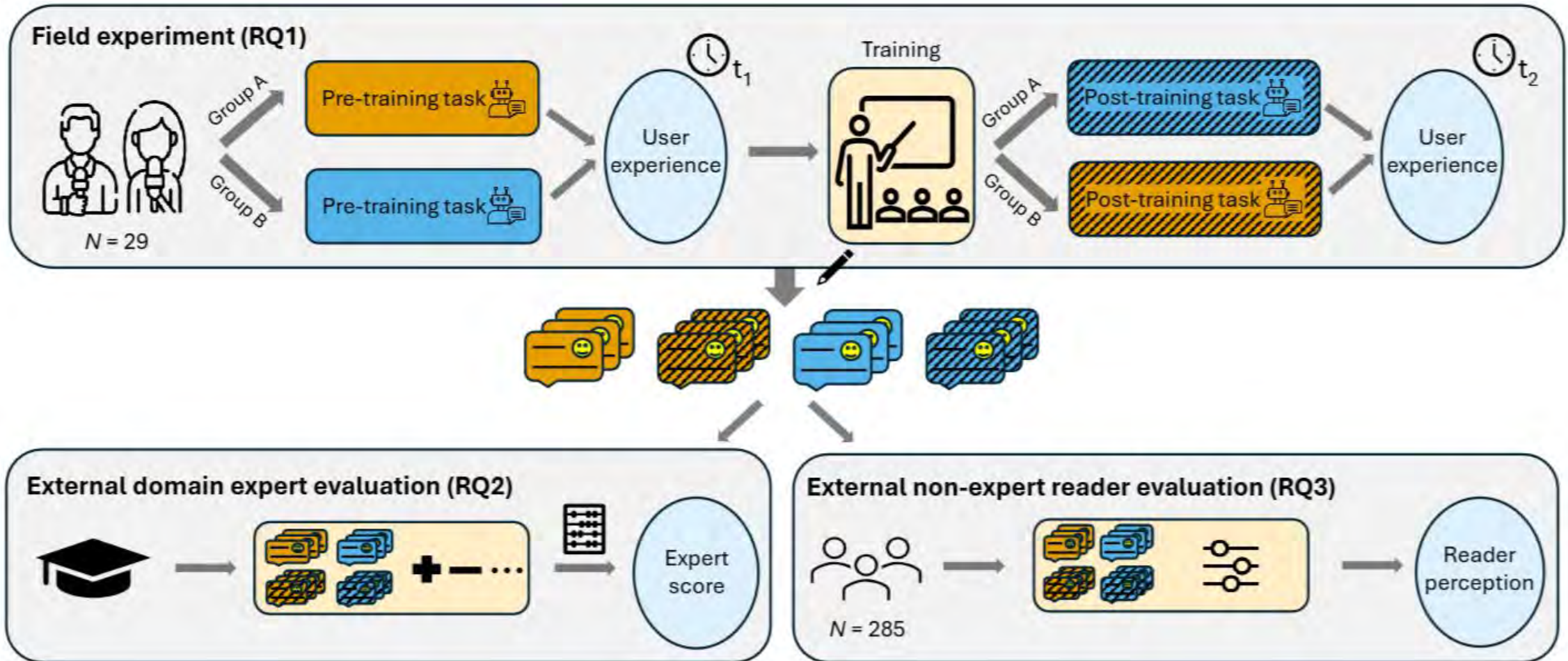


RQ 2: How does prompt engineering training influence the accuracy of text written by journalists with LLMs?



RQ3 How does prompt engineering training influence the non-expert reader perception of texts written by journalists with LLMs?

Research design



Gun Violence Exposure and Suicide Among Black Adults

Importance: Black individuals are disproportionately exposed to gun violence in the US. Suicide rates among Black US individuals have increased in recent years.

Objective: To evaluate whether gun violence exposures (GVEs) are associated with suicidal ideation and behaviors among Black adults.

Design, Setting, and Participants: This cross-sectional study used survey data collected from a nationally representative sample of self-identified Black or African American (hereafter, Black) adults in the US from April 12, 2023, through May 4, 2023.

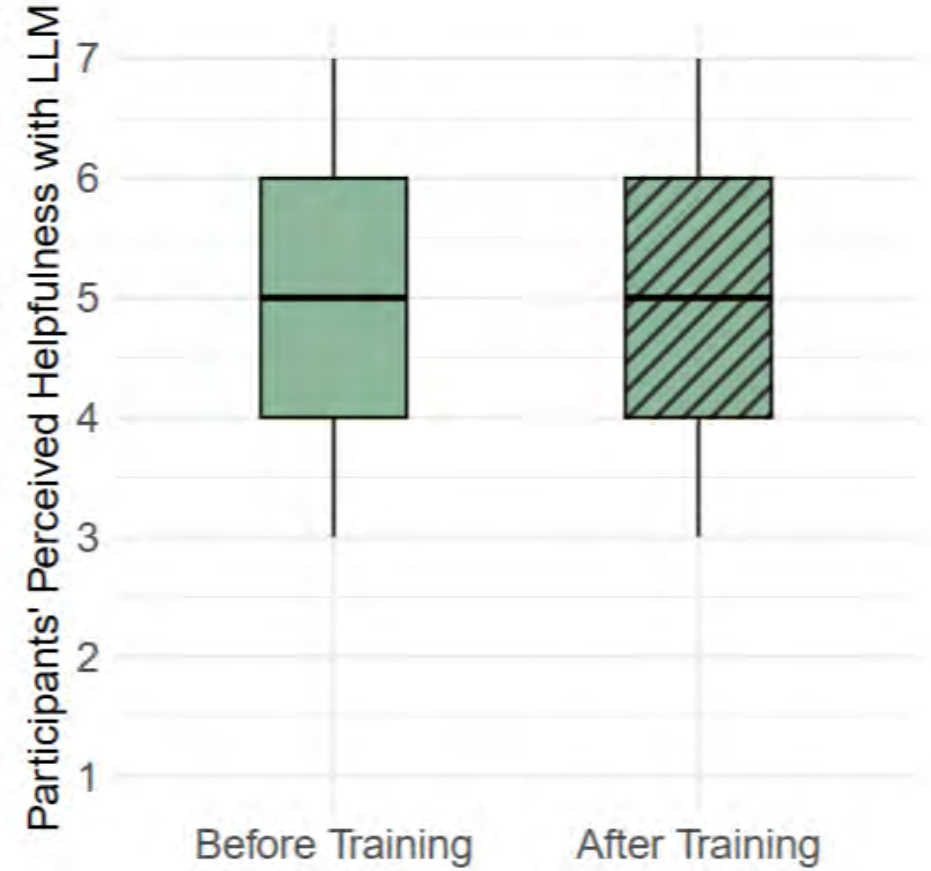
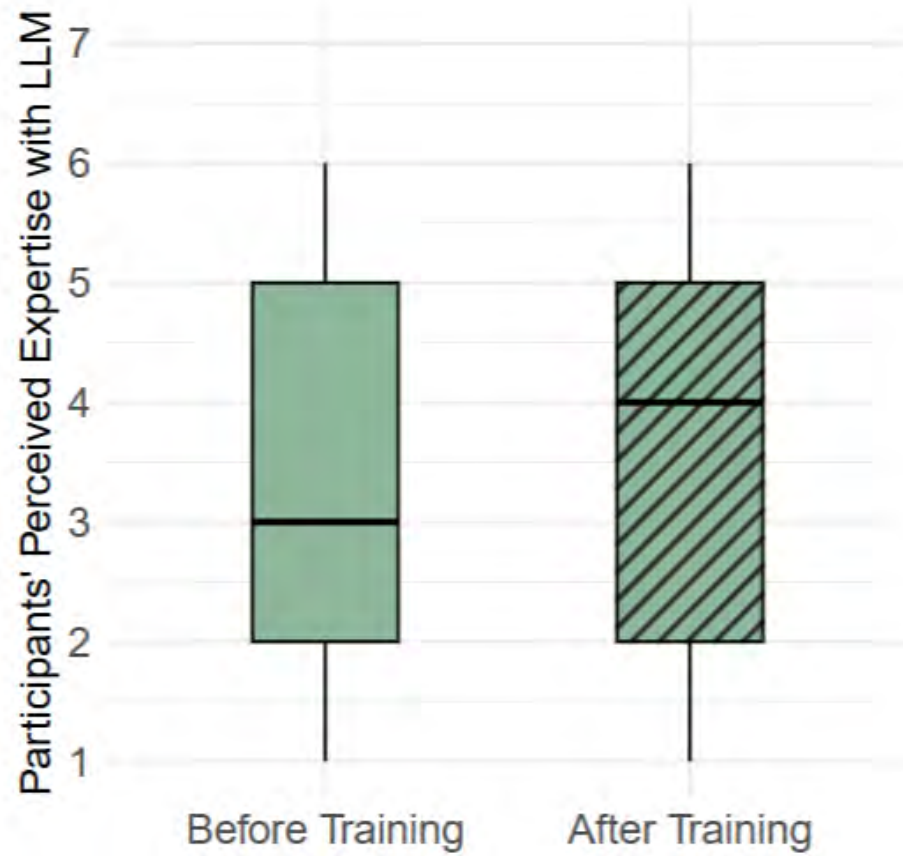
Exposures: Ever being shot, being threatened with a gun, knowing someone who has been shot, and witnessing or hearing about a shooting.

Main Outcomes and Measures: Outcome variables were derived from the Self-Injurious Thoughts and Behaviors Interview, including suicidal ideation, suicide attempt preparation, and suicide attempt. A subsample of those exhibiting suicidal ideation was used to assess for suicidal behaviors.

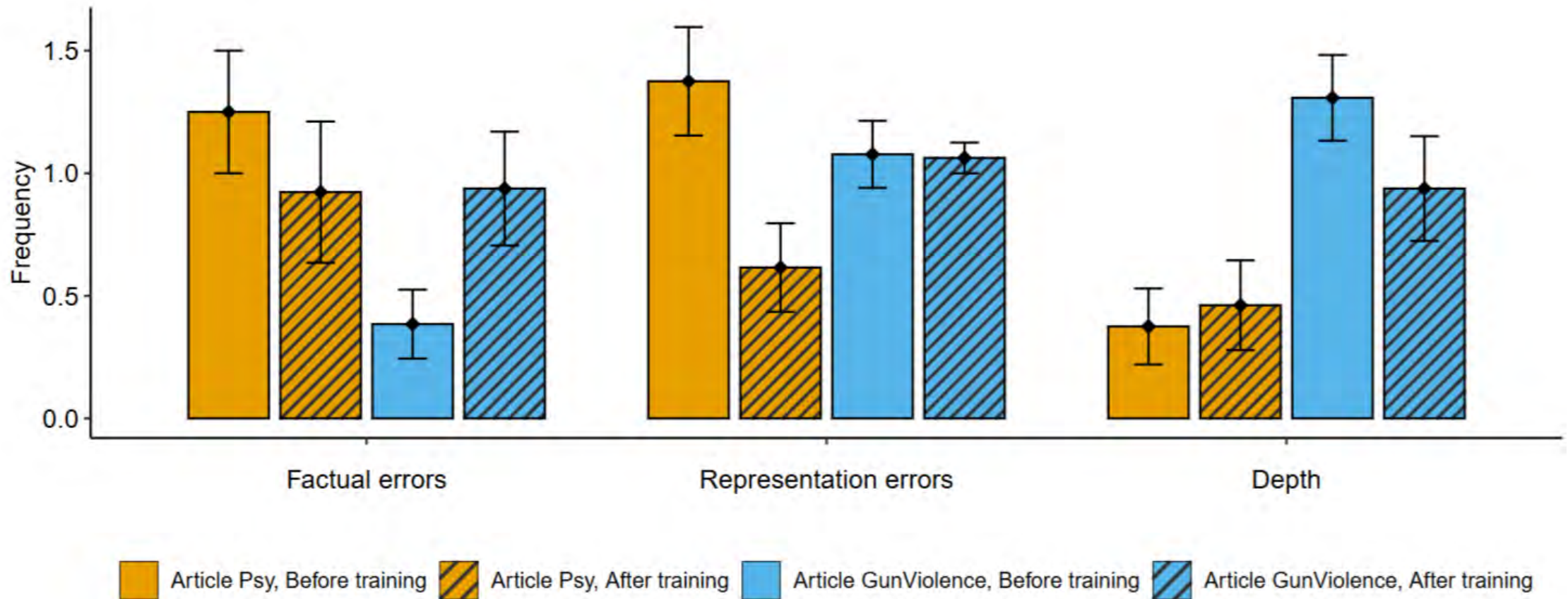
Results: The study sample included 3015 Black adults (1646 [55%] female; mean [SD] age, 46.34 [0.44] years [range, 18–94 years]). Most respondents were exposed to at least 1 type of gun violence (1693 [56%]), and 300 (12%) were exposed to at least 3 types of gun violence. Being threatened with a gun (odds ratio [OR], 1.44; 95% CI, 1.01–2.05) or knowing someone who has been shot (OR, 1.44; 95% CI, 1.05–1.97) was associated with reporting lifetime suicidal ideation. Being shot was associated with reporting ever planning a suicide (OR, 3.73; 95% CI, 1.10–12.64). Being threatened (OR, 2.41; 95% CI, 2.41–5.09) or knowing someone who has been shot (OR, 2.86; 95% CI, 1.42–5.74) was associated with reporting lifetime suicide attempts. Cumulative GVE was associated with reporting lifetime suicidal ideation (1 type: OR, 1.69 [95% CI, 1.19–2.39]; 2 types: OR, 1.69 [95% CI, 1.17–2.44]; ≥ 3 types: OR, 2.27 [95% CI, 1.48–3.48]), suicide attempt preparation (≥ 3 types: OR, 2.37; 95% CI, 2.37–5.63), and attempting suicide (2 types: OR, 4.78 [95% CI, 1.80–12.71]; ≥ 3 types: OR, 4.01 [95% CI, 1.41–11.44]).

Conclusions and Relevance: In this cross-sectional study, GVE among Black adults in the US was significantly associated with lifetime suicidal ideation and behavior. Public health efforts to substantially reduce interpersonal gun violence may yield additional benefits by decreasing suicide among Black individuals in the US.

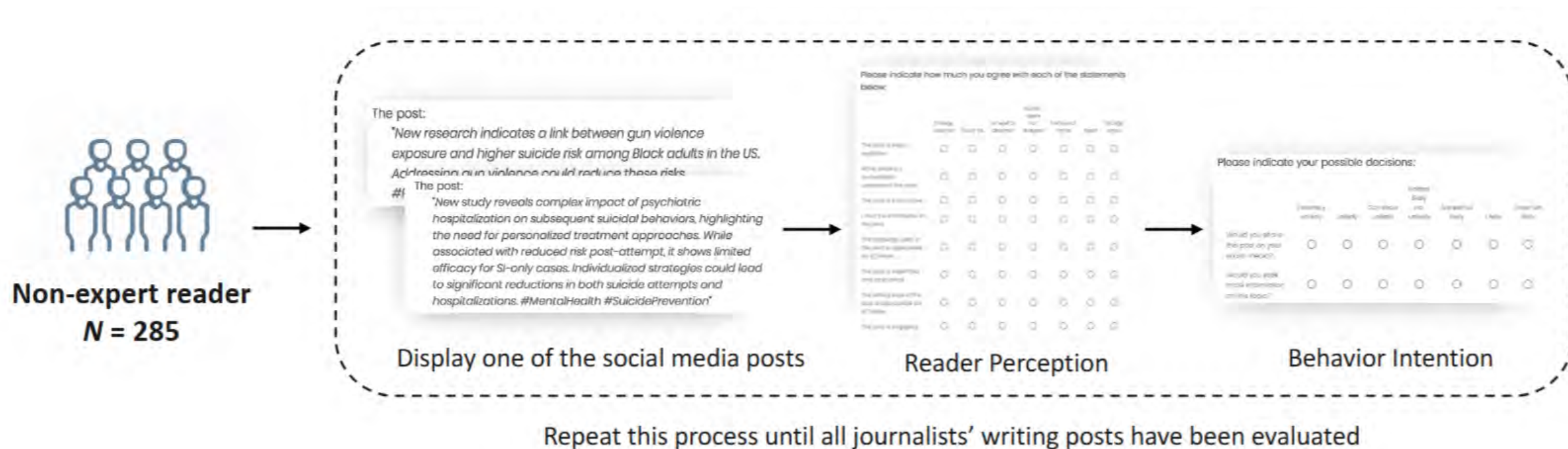
Perception of journalists before/after prompt education

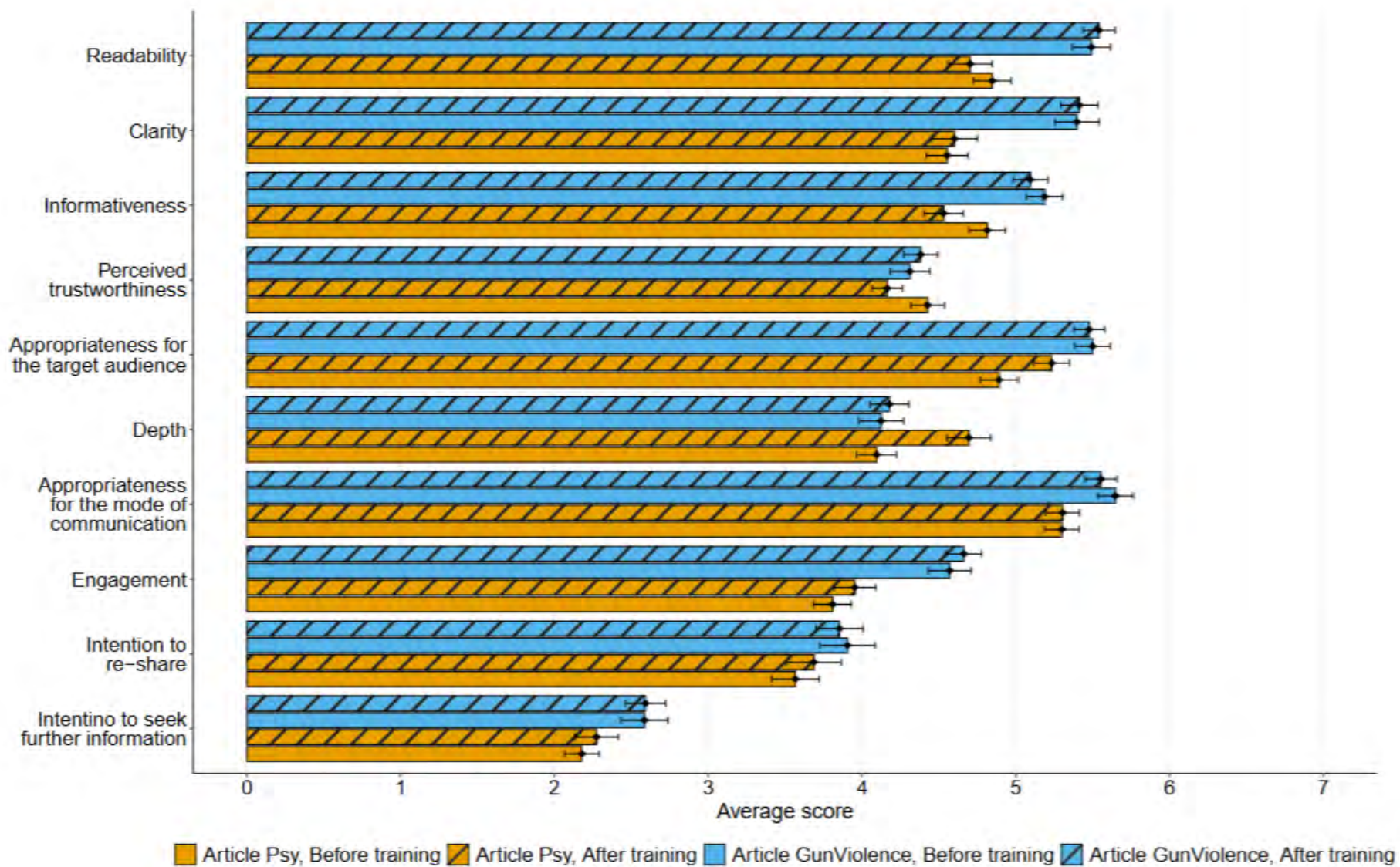


Accuracy of written social media posts (measured by domain experts)

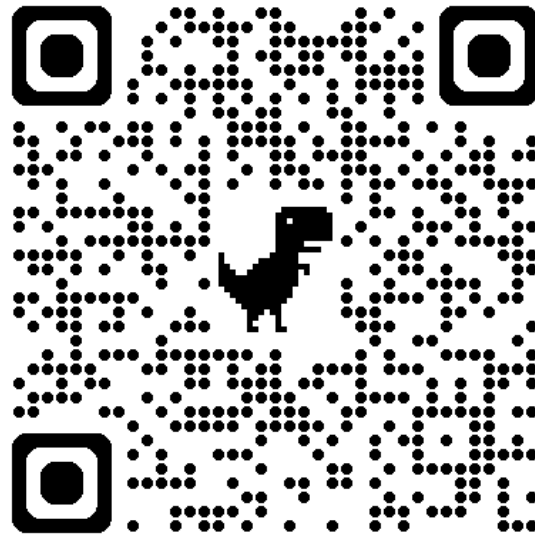


But what does “the reader” think?





Bashardoust, A., Feng, Y., Geissler, D., Feuerriegel, S., & Shrestha, Y. R. (2024). The Effect of Education in Prompt Engineering: Evidence from Journalists. *arXiv preprint arXiv:2409.12320*.



Optimal targeting

Whom to target?



How to target?

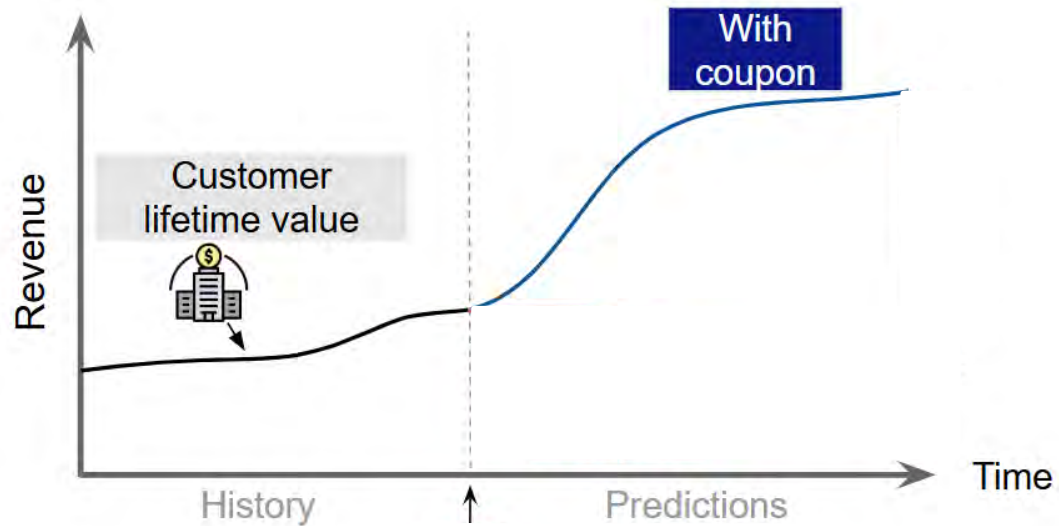


Climate change will harm your children

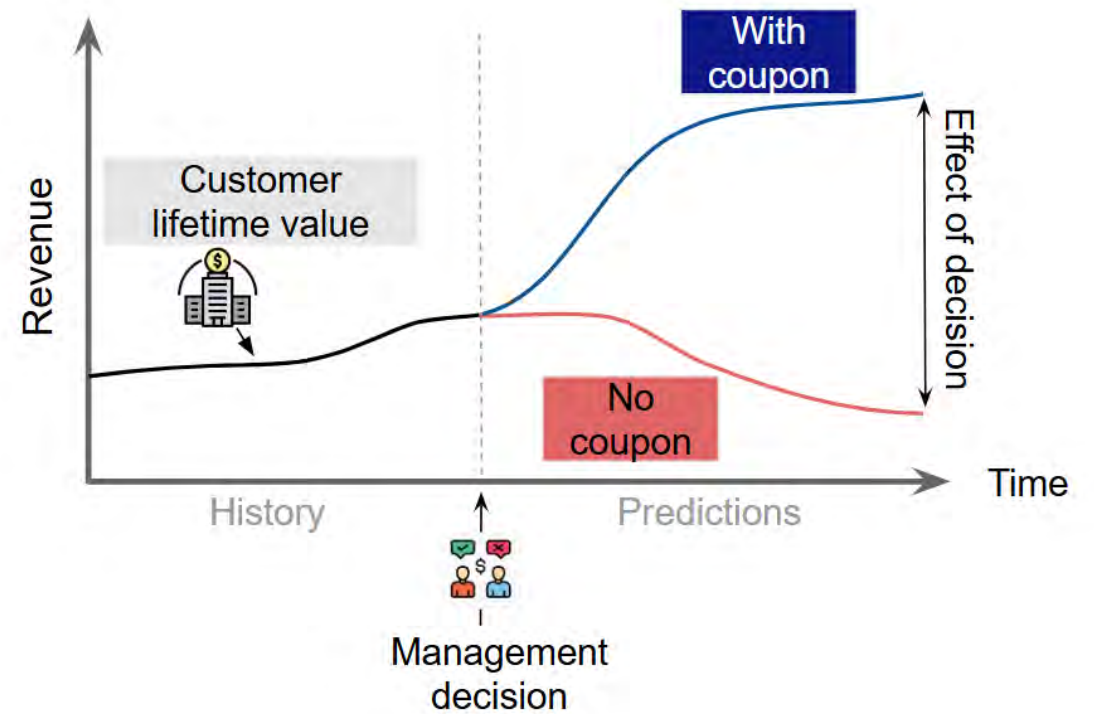
Climate change is a moral duty

Your friends fight climate change. Join them!

Traditional ML



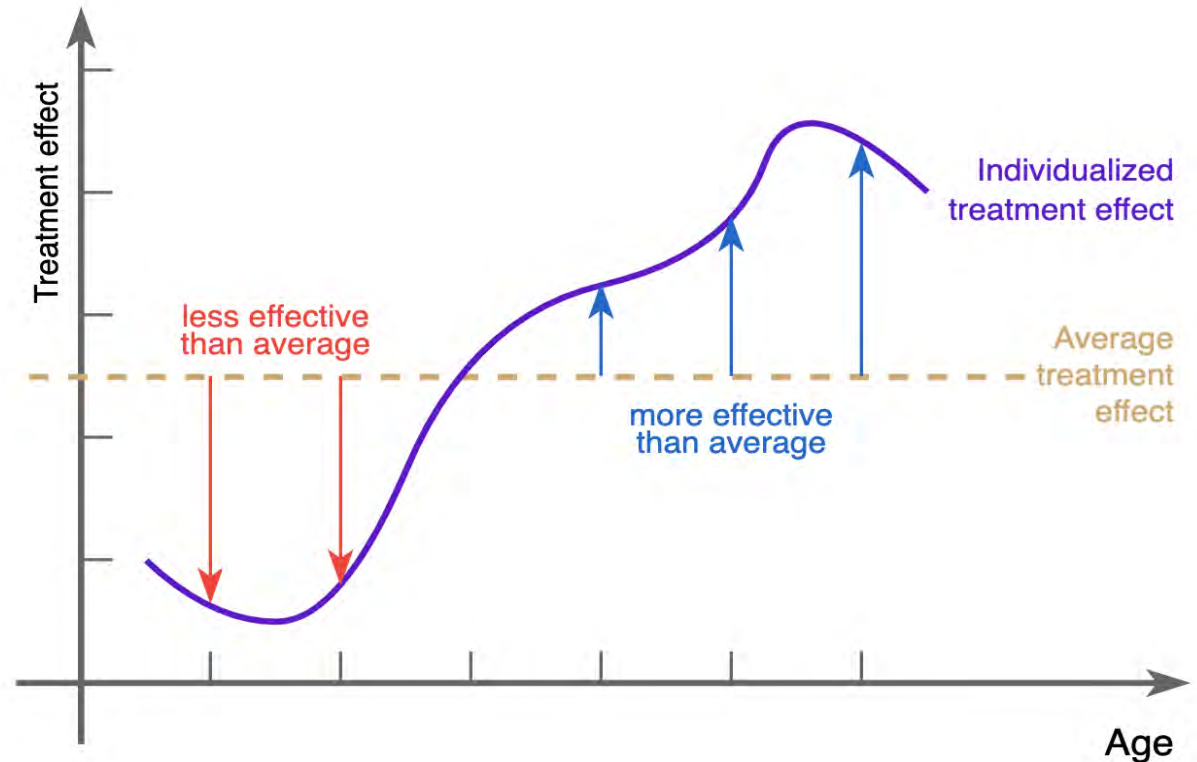
Causal ML



AIM

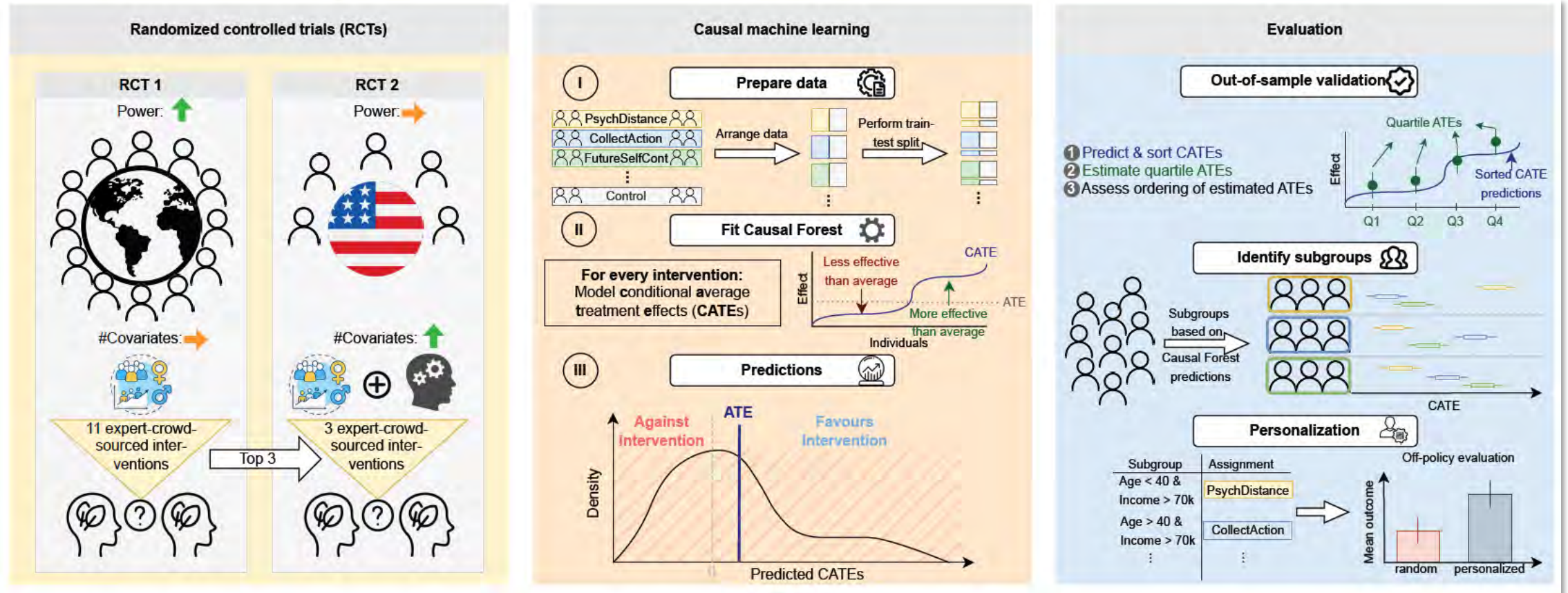
Understanding heterogeneity in the treatment effect

- Focus is often on **average** treatment effect (ATE)
- ATE is aggregated across the population
- ATE **cannot** tell whether a treatment works for some or not
→ e.g., medication works only for women but not for men, but RCT was done with all patients
- NB: both RCTs and target trial emulation focus on ATEs



To personalize treatment recommendations, we need to understand the **individualized** treatment effect (ITE)

Example: optimal nudging to improve climate beliefs



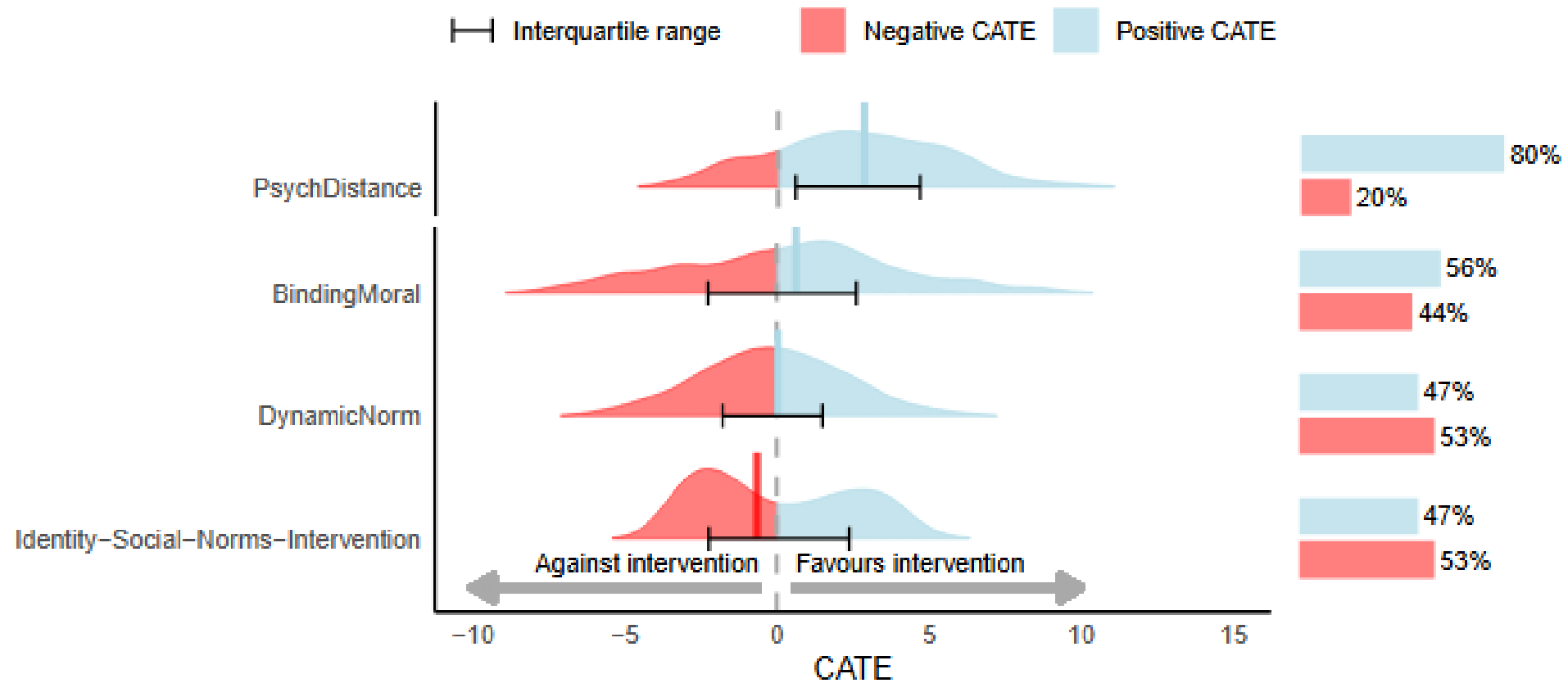
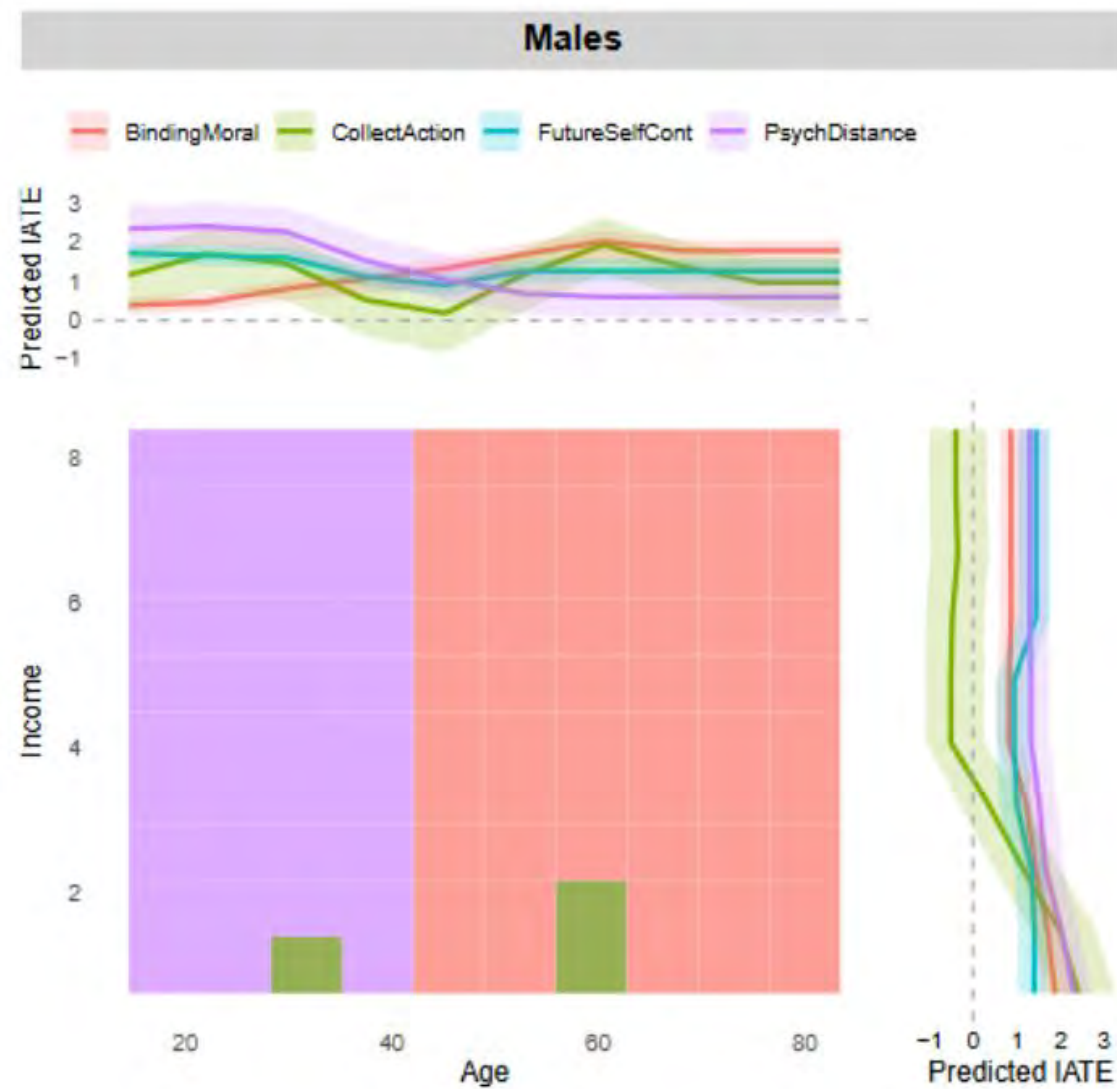
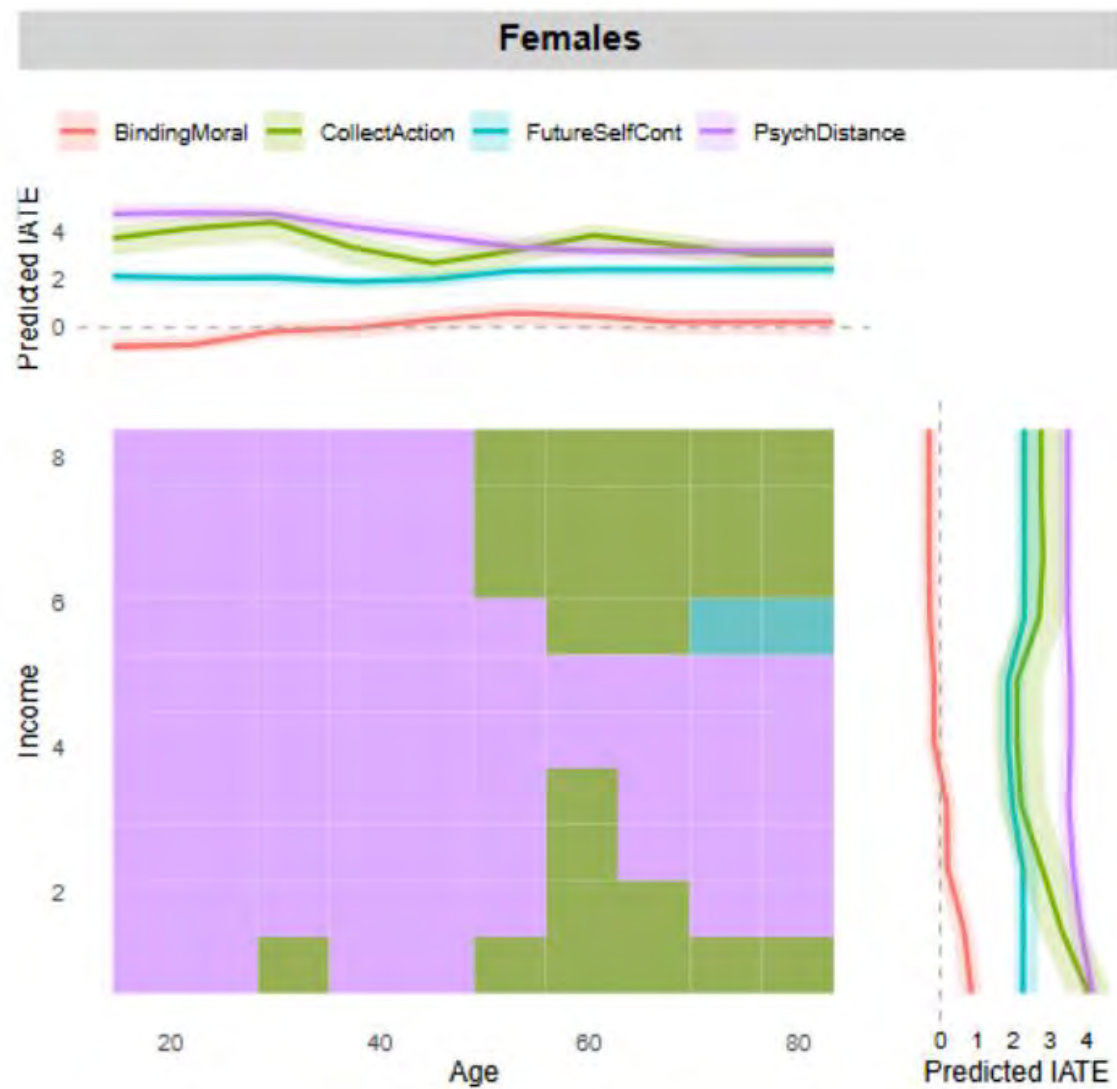


Figure 1: Distributions of the predicted conditional average treatment effects (CATEs) for all interventions. The interquartile range is plotted as black horizontal bars. The interventions are sorted by their average treatment effect (ATE) plotted as vertical bars.

Deriving optimal targeting rules





INSTITUTE OF AI IN MANAGEMENT



Institute of AI in Management
Prof. Dr. Stefan Feuerriegel

<http://www.ai.bwl.lmu.de>



@stfeuerriegel



stefan-feuerriegel

Artificial intelligence | Impact

Short introduction to causal machine learning




Reference:

Feuerriegel, S., Frauen, D., Melnychuk, V., Schweisthal, J., Hess, K., Curth, A., Bauer, S., Kilbertus, N., Kohane, I.S. and van der Schaar, M., 2024. Causal machine learning for predicting treatment outcomes. *Nature Medicine*, 30(4), pp.958-968.

Estimating the potential outcomes of treatments

Problem formulation

- Given i.i.d. observational dataset

 covariates
 (binary) treatments
 continuous (factual) outcomes

- We want to identify & estimate treatment outcomes:

- treatment effects











$$Y[1] - Y[0]$$






- potential outcomes

(separately) $Y[0]$ $Y[1]$

- Fundamental problem:**
never observing both potential outcomes!

$$\mathcal{D} = \{x_i, a_i, y_i\}_{i=1}^n \sim \mathbb{P}(X, A, Y)$$

Patient	Covariates 	Treatment 	Outcome  $Y = Y(0)$  $Y = Y(1)$	
		0	-1.0	
		1		2.3
		1		0.3
...

Patient	Covariates 	Potential outcomes $Y(0)$ $Y(1)$		Treatment effect $Y(1) - Y(0)$
		?	?	?
		?	?	?
...

A New Machine Learning Approach Answers ‘What If’ Questions

Causal ML enables managers to explore different options to improve decision-making.

By Stefan Feuerriegel, Yash Raj Shrestha, and Georg von Krogh

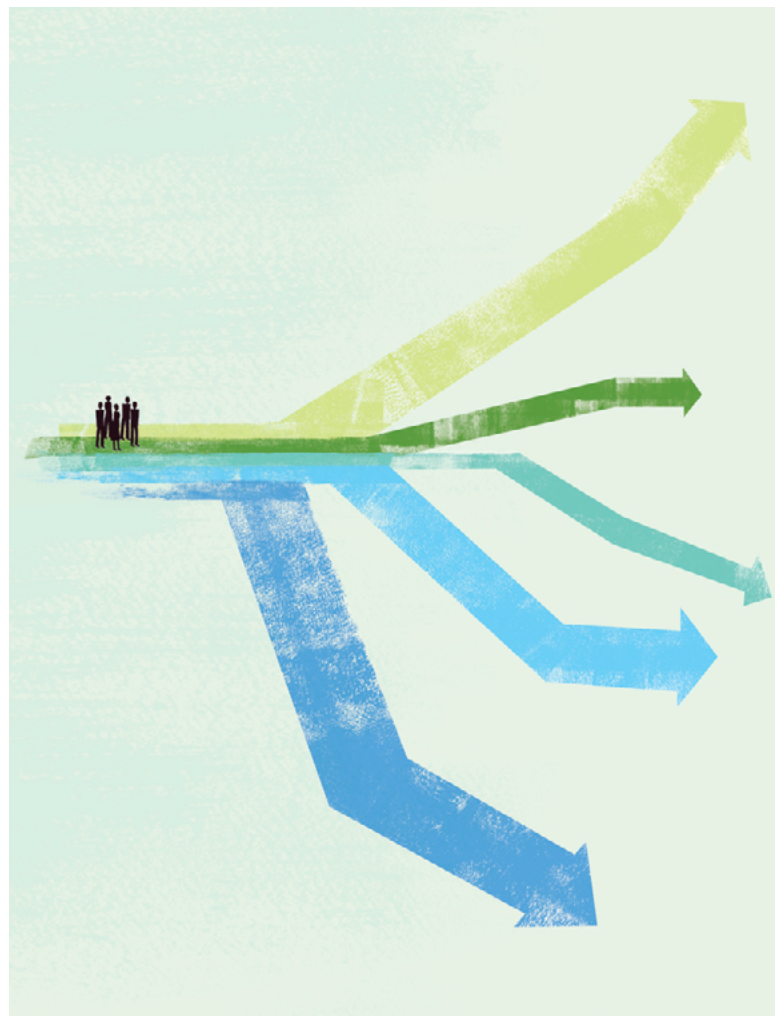
MACHINE LEARNING IS NOW widely used to guide decisions in processes where gauging the probability of a specific outcome — such as whether a customer will repay a loan — is sufficient. But because the technology, as traditionally applied, relies on correlations to make predictions, the insights it offers managers is flawed, at best, when it comes to anticipating the impact of different choices on business outcomes.¹

Consider leaders at a large company who must decide how much to invest in R&D in the coming year. Using traditional ML, they can ask what will happen when they increase their spending. They might find a strong correlation between higher levels of investment and higher revenue when the economy is growing. And they might conclude that, since economic conditions are favorable, they should increase the R&D budget.

But should they really? If so, by how much? External factors, such as levels of consumer spending, technology spillover from competitors, and interest rates, also influence revenue growth. Comparing how different levels of investment might affect revenue while considering these other variables is useful for the manager who is trying to determine the R&D budget that will deliver the greatest benefit to the company.

Causal ML — an emerging area of machine learning — can help to answer such what-if questions through causal inference. Similar to how marketers use A/B tests to infer which of two ads is likely to generate more sales, causal ML can inform what might happen if managers were to take a particular action.²

This makes the technology useful in many of the same business functions that use traditional ML, including product development, manufacturing, finance, human resources, and marketing.³ Traditional ML is still the go-to approach when the only goal is to make predictions — such as whether stock prices will rise or which products customers are most likely to buy. When a company wants



to predict what would happen if it were to make one decision versus another — such as whether a 10% discount or none is more likely to induce a customer to make a repeat purchase — it needs causal ML.

Our research on machine learning and AI and our experience helping companies apply causal ML points

out a path to using the technology successfully. (See “The Research.”) Companies will need the right expertise, too, and should boost employees’ literacy in causal ML.

What Causal ML Can and Cannot Do

Causal ML is a powerful tool, but managers may find the name misleading. The label “counterfactual prediction” would more accurately reflect what it does: predict outcomes based on hypothetical actions. The technology is best understood as a way to make better guesses rather than as a source of definitive answers. Framing it in this way can remind managers not to overinterpret the results.

It does this using causal inference, which looks at past results to understand cause-and-effect relationships among variables. But instead of focusing on why something happened, causal ML applies these relationships to predict the effects of interventions in new, forward-looking settings.

However, the method cannot explain why a causal relationship exists between a particular factor and the outcome it affects. For instance, a causal ML model might predict that reducing an R&D budget will decrease revenue, but it will not explain why that relationship exists or whether confounders — factors affecting both the decision and the outcome — might change and invalidate that prediction. Managers should use their domain expertise to evaluate whether a given prediction makes sense. This approach helps ensure that the model’s predictions are interpreted correctly and remain relevant to real-world decisions.

Like traditional machine learning, causal ML is most effective when managers have large volumes of data, their options are clearly defined, and the desired outcome is well understood. It is generally unsuitable for one-off decisions and in scenarios requiring intuition or creativity.

Choose the Right Problem — and Data

Causal ML is best at predicting the outcomes of straightforward decisions that are supported by ample historical data from internal and external sources. Questions about operations can be good candidates for the approach because they are made frequently and companies have a lot of data to support them.⁴ The following are examples of causal ML’s use in that context:

- Booking.com collects data from thousands of hotel reservations every hour. Marketers at the company use causal ML to determine not only whether to give discounts but also which customers should get them.
- Chocolate maker Lindt has extensive data about environmental conditions, equipment, packaging, and other factors that affect the quality of its

world-famous truffles. Manufacturing managers use causal ML to help them fine-tune parameters such as the temperature of machines and the configurations of truffle molds.⁵

- Hitachi ABB Power Grids turned to causal ML to reduce failure rates in its semiconductor manufacturing process, using machine performance data. It was able to cut its yield loss by about half by identifying which combination of machines consistently produced the best-quality chips.⁶

At Novartis, managers who had been educated about the capabilities of different kinds of machine learning were able to identify several decision-making tasks where replacing traditional machine learning with causal ML offered significant benefits. They had asked a traditional ML model whether increasing the marketing budget would increase sales, but its predictions were not helping them decide how to allocate that budget. They decided to use causal ML to evaluate how different promotional campaigns might affect future sales. They used the predictions to distribute resources to the campaigns that were likely to be most effective.

A decision that is suitable for causal ML can be expressed as a number or a binary choice (such as an amount of revenue or buy/hold). It may also be framed as a question about which action to take: to allocate a marketing budget of \$10,000 or \$15,000 for the next quarter, or to offer a 10% discount or none on a product.⁷

Further, causal ML cannot effectively address every potential use case, even if it seems suitable for that on the surface. Confounders — the variables that affect both the outcome and the decision — introduce biases that affect predictions and must be accounted for. They can be challenging or impossible to test for, and they affect the accuracy of predictions. If, for example, data is available only for product sales during an economic upturn, predictions of product sales during a downturn would be less reliable.

When managers have determined what they want to decide, identified how they will measure the outcome, and affirmed that they have enough data, they can begin to work with data scientists to assemble and categorize that data to build their causal ML model. Business leaders and other individuals with domain knowledge are essential partners to data scientists and machine learning experts in building causal ML models that provide reliable results.

Training the model to capture complex cause-and-effect relationships requires data from at least a few dozen — and ideally, hundreds or thousands — of historical decisions. With massive amounts of data, the model can uncover connections between variables that may be unknown to managers or difficult to quantify. Less data

leads to less-accurate predictions.

Broadly, causal ML requires three categories of data that were alluded to above: decisions, outcomes, and confounders. Decision data encompasses what managers have done in the past, such as the staffing levels or budgets they set, discounts they offered, investments they made, or processes they changed. Outcome data may include any measurable business result, such as sales volume, revenue growth, quality metrics, or productivity.

Confounders can come from internal or external sources. They may include economic conditions, workforce composition, and competitor behavior, and they can vary with the decision being made. For a marketing decision, the type of device customers use may be a confounder because those with more expensive smartphones may tend to spend more money whether or not they respond to an incentive.

For example, *Neue Zürcher Zeitung*, an international media company that publishes the largest-circulation newspaper in Switzerland, implemented causal ML to improve the effectiveness of editors' content promotion decisions. The decision variable was whether an online article was promoted on one of two front pages that were served to readers. The outcome variable was a performance score that combined website traffic, reader engagement, and subscription signups. Confounders included time factors (such as the hour of the day), content characteristics (such as the article format), past performance indicators (including clicks), and past promotion decisions (including whether the article had been promoted elsewhere).

Identify Possible Causal Factors

A valuable lesson from our work has been the value of sketching a causal graph on a whiteboard that illustrates the expected relationships between the outcome, the decision, and the confounders at the start of the model development process. Managers' knowledge and expertise are essential here because they have repeatedly made decisions and learned to anticipate certain results.

The causal graph tells the data scientists (who should be experts in causal inference) whether to treat a variable as a cause or an effect in the model. In this way, the team can rule out reverse causality errors. That is, they can ensure that the model does not misinterpret one variable as causing another when, in reality, the effect is the opposite.

Imagine a celebrity with millions of social media followers. If we do not know much about social media or stardom, we might conclude that fame comes from having a high follower count. The reverse is more likely to be true. As even the average teenager has observed, to get

millions of strangers to follow their social media accounts, they first have to do something that gets them noticed.

In the case of our R&D spending question, the budget influences revenue, not the other way around. Meanwhile, confounders such as the economic climate, market trends, or team expertise are acknowledged as driving both the budget decision and business outcomes but are not influenced by either. The model would take all of this into account. (See "A Causal Graph for an R&D Budget Decision.")

Choose the Output

Next, managers need to choose the type of answer the model should give in response to the question (referred to in statistics as the output, or *estimand*): It can predict the end result of a decision or the relative benefit of one alternative compared with another.

Each of those outputs can be useful, depending on how the manager is thinking through a decision. Focusing on end results, such as potential revenues under different budget scenarios or personalized incentives for individual customers, helps with strategic planning. However, comparing the incremental effects of different decisions is often sufficient for making one: If a manager wants to know which of two ads is likely to boost sales more effectively, they do not necessarily need to predict how much revenue each variant might generate. They only need to know the relative benefit: that one ad is likely to generate three times more revenue than the other. Moreover, focusing on relative benefits generates more reliable predictions than focusing on end results. We recommend pursuing only as much granularity as is necessary.

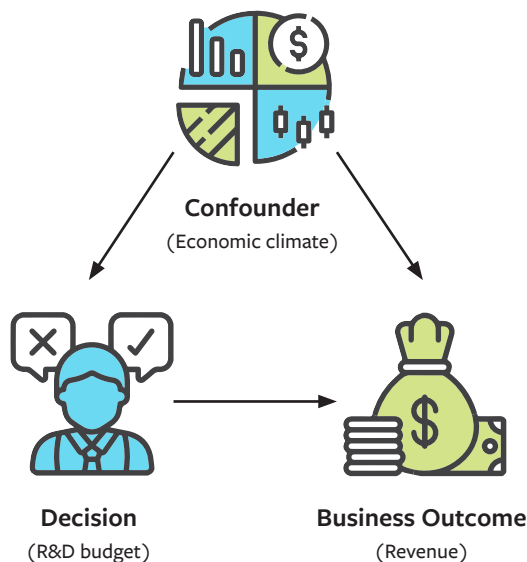
Editors from *Neue Zürcher Zeitung* were interested in predicting the actual click rates for each article they promoted, but the company opted instead to predict the likely net gain in performance from promoting an item. This approach enabled causal ML to make more accurate predictions about which content, when promoted, would increase clicks and subscriptions. Editors learned that promoting articles written by the editor in chief significantly increased both outcomes.⁸ They had been promoting the top editor's articles sparingly, and the findings served as a starting point to revise their promotional strategy.

THE RESEARCH

- The authors worked with companies in manufacturing, pharmaceuticals, finance, travel, and media to develop and implement causal ML models for a variety of business functions. The companies included Booking.com, EthonAI, Hitachi ABB Power Grids, Neue, Novartis, UBS, and *Zürcher Zeitung*.
- Over five years, they implemented and evaluated the models and documented the practices that contributed to the companies' success using the technology.

A Causal Graph for an R&D Budget Decision

A causal graph can help data scientists understand whether a variable should be treated as a cause or an effect in a causal ML model. These graphs describe the role of different variables and how they are expected to interact. The arrows indicate the direction of influence for each. Here, the decision is the R&D budget allocation, which influences the business outcome (revenue). The economic climate is a confounder that influences both the decision and the outcome.



Train, Test, and Validate the Model

Once managers have defined the decision they want to make and their preferred type of output, data and machine learning scientists can choose the causal ML model that is right for the job. Once the model is implemented, machine learning engineers will train it using the previously categorized data.

The final step is to test and validate the causal ML model in practice to ensure that it is reliable and that its predictions result in better business performance. Validation also offers the opportunity for decision makers, including senior leaders, to gain trust in its predictions. Starting with relatively simple and straightforward problems where clear decision alternatives can be identified and assessed makes this step easier to accomplish.

Testing and validation require care because managers can observe only the outcome of the decision that was made in the real world. They have no way to know what the outcome would have been had a different decision

been made. Two strategies, “human in the loop” and the familiar A/B testing approach, have proved successful.

Neue Zürcher Zeitung chose to integrate the model’s recommendations with human decision-making processes.⁹ Its causal ML model recommends which content to promote, but the editors make the final decisions. The model relies on the same information that editors previously used to make their promotion decisions, so they can trust that the model is not missing key elements. The causal ML model’s recommendations typically match the editors’ own gut feelings, which gives them confidence that it is reliable.

Some decisions are tricky, and editors know that their judgment is not perfect. In cases where causal ML recommends a different decision than they would have made, the editors can test the recommendation and see the result. Over time, they should see that causal ML is able to make reliable recommendations in ambiguous situations. At that point, they will be able to follow the causal ML recommendations instead of their instincts more frequently.

Hitachi ABB used A/B testing to validate the causal ML models it built to improve manufacturing quality. In one application, managers used the model to predict which of several machines would produce the best-quality output in the etching and implantation steps of the semiconductor fabrication process and contribute to the highest-quality output overall.

To confirm that the predictions were reliable, managers did a controlled experiment in which they changed the machine used for etching and implantation and kept the machines used for other processes the same. They found that the better machine for etching and implantation was the same one that the causal ML model had predicted. Thanks to causal ML, managers were able to find and address the source of manufacturing issues more efficiently than they could have with either manual methods or traditional ML.¹⁰

Prepare the Organization

While causal ML has the potential to improve decisions, implementing such systems requires a high level of AI literacy in the workforce, specialized technical expertise, and patience — because these projects may take longer to develop than traditional ML applications. Managers can prepare their organizations by educating themselves and their workforces about causal AI and building the interdisciplinary teams needed to develop the applications.

Many companies today are investing heavily in educating employees about traditional ML and generative AI models (such as ChatGPT) to stay competitive and innovative. If the organization plans to use causal ML, it

Data scientists and ML engineers need to work closely to develop and implement causal ML.

needs to include this technology in its AI literacy efforts. Employees who are alert to the strengths and limitations of different AI approaches will be empowered to find opportunities to use them effectively.

We found that to excel at using causal ML, teams need strong expertise in data science and machine learning, along with domain knowledge. However, building such teams can be costly, particularly when it requires companies to hire data scientists or turn to external consultants and partners.

Moreover, data scientists and machine learning engineers are typically assigned to different teams. They need to work closely when developing and implementing causal ML models and have strong engagement with the business stakeholders who have domain knowledge. (Domain knowledge is also essential in traditional machine learning but is often less rigorously applied because teams do not deeply consider the underlying relationships between variables when building those models.)

For example, at *Neue Zürcher Zeitung*, the insights that editors and marketers have into editorial processes, customer preferences, and the long-term objectives of the brand help data scientists define variables that measure those factors. At Hitachi ABB, engineers supplied the insight to define which production variables to include in the models.

Interdisciplinary teams are often plagued by a lack of common understanding, vocabulary, and ways of working. Managers need to foster an environment where cross-functional collaboration can thrive and all relevant stakeholders are involved throughout the model development process. Regular workshops, meetings, and training sessions where data scientists, machine learning engineers, and domain experts jointly explore problems, refine models, and discuss the implications of the findings together can foster an environment in which cross-functional collaboration thrives.

MACHINE LEARNING HAS CHANGED HOW numerous organizations make decisions; causal ML can deepen insights further by predicting the effects of different choices on business outcomes. Companies are more likely to benefit from machine learning when decision makers trust the results. Knowing what causal ML can do and how it compares with traditional ML can help

them choose the right projects for each technology and increase their success rates.

When managers use causal ML prudently to explore the options for straightforward decisions, they can significantly improve their operations — and, ultimately, their financial results. ■

Stefan Feuerriegel is the director of the Institute of AI in Management in the LMU Munich School of Management. **Yash Raj Shrestha** is the group head at the Applied Artificial Intelligence Lab at the University of Lausanne. **Georg von Krogh** is a professor and chair of strategic management and innovation at ETH Zurich and an associated faculty member at the ETH AI Center.

REFERENCES

1. S. Feuerriegel, Y.R. Shrestha, G. von Krogh, et al., “Bringing Artificial Intelligence to Business Management,” *Nature Machine Intelligence* 4, no. 7 (July 2022): 611-613; and P. Hünemund, J. Kaminski, and C. Schmitt, “Causal Machine Learning and Business Decision-Making,” SSRN, updated Feb. 19, 2022, <https://ssrn.com>.
2. S. Feuerriegel, D. Frauen, V. Melnychuk, et al., “Causal Machine Learning for Predicting Treatment Outcomes,” *Nature Medicine* 30 (April 2024): 958-968; V. Chernozhukov, C. Hansen, N. Kallus, et al., “Applied Causal Inference Powered by ML and AI” (pub. by the authors, July 2024), PDF; and C. Fernández-Loría and F. Provost, “Causal Decision-Making and Causal Effect Estimation Are Not the Same ... and Why It Matters,” *Informa Journal on Data Science* 1, no. 1 (April-June 2022): 4-16.
3. M. von Zahn, K. Bauer, C. Mihale-Wilson, et al., “Smart Green Nudging: Reducing Product Returns Through Digital Footprints and Causal Machine Learning,” *Marketing Science, Articles in Advance*, published online Aug. 8, 2024; E. Ascarza, “Retention Futility: Targeting High-Risk Customers Might Be Ineffective,” *Journal of Marketing Research* 55, no. 1 (February 2018): 80-98; J. Yang, D. Eckles, P. Dhillon, et al., “Targeting for Long-Term Outcomes,” *Management Science* 70, no. 6 (June 2024): 3841-3855; and M. Kraus, S. Feuerriegel, and M. Saar-Tsechansky, “Data-Driven Allocation of Preventive Care With Application to Diabetes Mellitus Type II,” *Manufacturing & Service Operations Management* 26, no. 1 (January-February 2024): 137-153.
4. G. von Krogh, S.M. Ben-Menahem, and Y.R. Shrestha, “Artificial Intelligence in Strategizing: Prospects and Challenges,” in “Strategic Management: State of the Field and Its Future,” eds. I.M. Duhaime, M.A. Hitt, and M.A. Lyles. (New York: Oxford University Press, 2021), 625-646.
5. “Premium Chocolate Production Perfected: AI’s Role in Quality Excellence,” ETH AI Center, Dec. 11, 2023, <https://ai.ethz.ch>.
6. J. Senoner, T. Netland, and S. Feuerriegel, “Using Explainable Artificial Intelligence to Improve Process Quality: Evidence From Semiconductor Manufacturing,” *Management Science* 68, no. 8 (August 2022): 5704-5723.
7. H. Wasserbacher and M. Spindler, “Machine Learning for Financial Forecasting, Planning and Analysis: Recent Developments and Pitfalls,” *Digital Finance* 4 (March 2022): 63-88.
8. J. Persson, S. Feuerriegel, and C. Kadar, “Off-Policy Learning for Audience-Wide Content Promotions,” working paper, 2023.
9. Ibid.
10. Senoner et al., “Using Unexplainable Artificial Intelligence,” 5704-5723.

Reprint 66336. For ordering information, see page 4. Copyright © Massachusetts Institute of Technology, 2025. All rights reserved.

Causal machine learning for predicting treatment outcomes

Received: 3 January 2024

Accepted: 4 March 2024

Published online: 19 April 2024



Stefan Feuerriegel^{1,2}✉, **Dennis Frauen**^{1,2}, **Valentyn Melnychuk**^{1,2},
Jonas Schweisthal^{1,2}, **Konstantin Hess**^{1,2}, **Alicia Curth**³, **Stefan Bauer**^{4,5},
Niki Kilbertus^{2,4,5}, **Isaac S. Kohane**⁶ & **Mihaela van der Schaar**^{7,8}

Causal machine learning (ML) offers flexible, data-driven methods for predicting treatment outcomes including efficacy and toxicity, thereby supporting the assessment and safety of drugs. A key benefit of causal ML is that it allows for estimating individualized treatment effects, so that clinical decision-making can be personalized to individual patient profiles. Causal ML can be used in combination with both clinical trial data and real-world data, such as clinical registries and electronic health records, but caution is needed to avoid biased or incorrect predictions. In this Perspective, we discuss the benefits of causal ML (relative to traditional statistical or ML approaches) and outline the key components and steps. Finally, we provide recommendations for the reliable use of causal ML and effective translation into the clinic.

Assessing the effectiveness of treatments is crucial to ensure patient safety and personalize patient care. Recent innovations in ML offer new, data-driven methods to estimate treatment effects from data. This branch in ML is commonly referred to as causal ML as it aims to predict a causal quantity, namely, changes in patient outcomes due to treatment¹. Causal ML can be used to estimate treatment effects from both experimental data obtained through randomized controlled trials (RCTs) and observational data obtained from clinical registries, electronic health records and other real-world data (RWD) sources to generate clinical evidence. A key strength of causal ML is that it enables estimation of individualized treatment effects, as well as personalized predictions of potential patient outcomes (for example, survival, readmission, quality of life or toxicity) under different treatment scenarios. This offers a granular understanding of when treatments are effective or harmful, so that decision-making in patient care can be personalized to individual patient profiles. Still, cautious use is important as causal inference rests on formal assumptions that cannot be tested.

In this Perspective, we explain how causal ML differs from traditional statistical and ML approaches, and we discuss the essential components and steps for its use in the clinic. We provide recommendations

for avoiding common technical pitfalls and outline a path to translation of this approach into clinical practice.

Causal ML in medicine

In medicine, causal ML offers several opportunities for estimating individualized treatment effects from data, which eventually help in greater personalization of care. First, at the patient level, causal ML can handle high-dimensional and unstructured data with patient covariates, and thus estimate treatment effects from multimodal datasets containing images, text or time series, as well as genetic data. For example, one could estimate treatment effects from computed tomography scans or entire electronic health records. Second, at the outcome level, causal ML can help make personalized estimates of treatment effects for subpopulations or even predict outcomes for individual patients². For example, individual differences in drug metabolism can lead to serious side effects for drugs in some patients but can be lifesaving in others³, so a causal ML approach could learn such differences and thus help in designing personalized treatment strategies. Third, at the treatment level, causal ML can be effective for estimating heterogeneity in treatment effects across patients in a data-driven manner, to identify for which patient subgroups treatment is effective (Fig. 1c).

¹LMU Munich, Munich, Germany. ²Munich Center for Machine Learning, Munich, Germany. ³Department of Applied Mathematics & Theoretical Physics, University of Cambridge, Cambridge, UK. ⁴School of Computation, Information and Technology, TU Munich, Munich, Germany. ⁵Helmholtz Munich, Munich, Germany. ⁶Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. ⁷Cambridge Centre for AI in Medicine, University of Cambridge, Cambridge, UK. ⁸The Alan Turing Institute, London, UK. ✉e-mail: feuerriegel@lmu.de

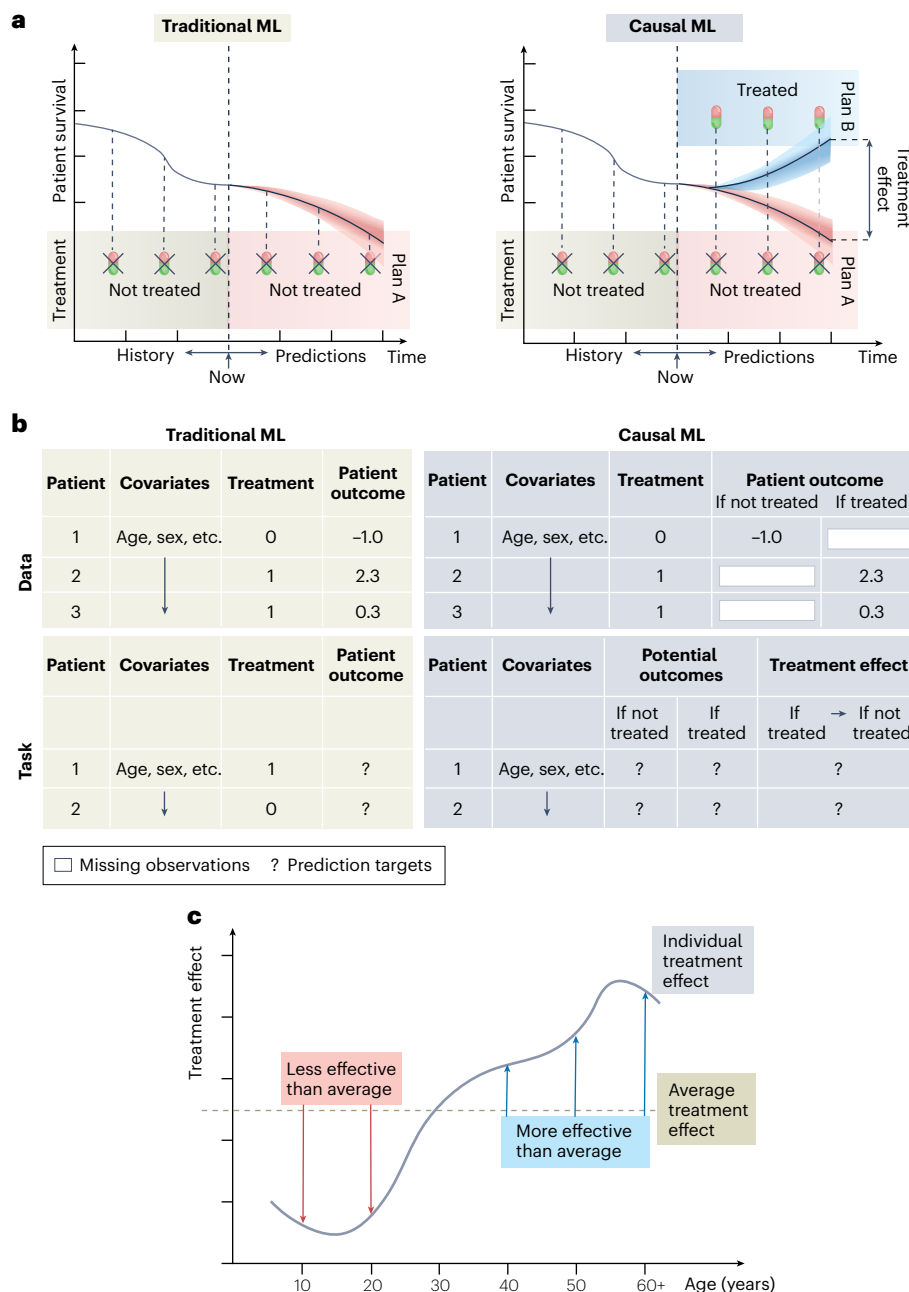


Fig. 1 | Causal ML for predicting treatment outcomes. **a**, Different from traditional ML, causal ML aims to (i) estimate the treatment effect or (ii) predict the potential patient outcomes themselves owing to treatments. As such, one can perform ‘what if’ reasoning to evaluate how patient outcomes will change due to administering a treatment. **b**, Causal ML is challenging owing to the ‘fundamental problem’ of causal inference—in that not all potential outcomes can be observed

and are thus missing in the data. Unless potential outcomes are explicitly needed, treatment effect estimation is preferred. **c**, Treatment effect heterogeneity refers to the variation in the response to treatment across different subgroups of a patient population (for example, according to age), indicating that the effectiveness of the intervention is not uniform for all individuals. For this, one must move beyond the ATE and obtain individualized treatment effects.

Despite these potential benefits, causal ML poses distinct challenges that necessitate custom methods. In addition, the appropriate application of this approach requires an understanding of how causal ML differs from traditional statistical and ML approaches.

When should I use causal ML?

Causal ML for estimating treatment effects is different from traditional predictive ML (see Box 1 for a glossary of terms). Intuitively, traditional ML aims at predicting outcomes⁴, while causal ML quantifies changes in outcomes due to treatment, so that treatment effects can be estimated (Fig. 1a). A typical use case for traditional ML is risk scoring, such as predicting the probability of diabetes onset to understand which patients

are at high risk—but without saying what the best treatment plan is^{5–9}. By contrast, causal ML aims to answer ‘what if’ questions. For example, causal ML could estimate how the risk of diabetes onset will change if the patient receives an antidiabetic drug^{10–12}, so that decisions can be made about whether to administer such a drug. Causal ML can also be used to predict the potential patient outcomes in response to different treatments. For example, in oncology, causal ML could make individualized predictions of survival under different treatment plans, which can then help medical practitioners in choosing a treatment plan that promises the largest chance of survival or longest duration of survival¹³.

Methods for estimating treatment effects have a long tradition in the statistical literature (for example, refs. 14–17). Causal ML builds

BOX 1

Glossary of common terms in causal ML

Causal graph: A graphical representation of the causal relationships between variables, typically using directed acyclic graphs to depict causal paths.

Causal ML: A branch of ML that aims to estimate causal quantities (for example, ATE and CATE) or to predict potential outcomes. Here, 'causal' implies that the target is a causal quantity when certain assumptions about the data-generating mechanism are satisfied. For alternative definitions and use cases of causal ML, see ref. 1.

Confounder: A variable that influences both the treatment assignment and the outcome.

Consistency: The potential outcome is equal to the observed patient outcome under the selected treatment, which implies that the potential outcomes are clearly defined and observable in principle.

Counterfactual outcome: The unobservable patient outcome that would have occurred, had a patient received a different treatment.

Factual outcome: The observed patient outcome that occurred for the observed treatment.

Identifiability: A statistical concept referring to the ability of causal quantities such as treatment effects to be uniquely inferred from the observed data.

Positivity: Each patient has a bigger-than-zero probability of receiving/not receiving a treatment. This is also called overlap assumption.

Potential outcome: The hypothetical patient outcome that would be observed if a certain treatment was administered.

Propensity score: The propensity score is the probability of receiving the treatment given the observed specific patient characteristics.

SUTVA: The outcome for any patient does not depend on the treatment assignment of other patients, and there is no hidden variation in the effect of the treatment across different settings or populations.

Unconfoundedness: Given observed covariates, the treatment assignment is independent of the potential outcomes. This is the case, for example, when there are no unobserved confounders, that is, variables influencing both the treatment and the outcome. The assumption is also called ignorability.

upon the same problem setup but makes changes to the estimation strategy. Hence, the core benefit of using causal ML is generally not the types of questions that can be asked, but how these questions can be answered. As such, causal ML can have benefits over alternative methods from the statistical literature (Box 2). First, methods from

BOX 2

Comparison of causal ML versus traditional statistics

Owing to the importance of treatment effect estimation across many application areas, methods for treatment effect estimation have been developed in different disciplines, including statistics, biostatistics, econometrics and ML (for example, refs. 23,27,49,53,54,61,98,99). However, there is no 'dichotomy' as many concepts are shared across disciplines. For example, many state-of-the-art methods for estimating treatment effects are model-agnostic in that they can be used in combination with both arbitrary models from classical statistics and also more modern ML models^{23,49,61}.

Eventually, the choice of whether to rely on a classical statistical model or a more modern ML method presents a trade-off that depends on the underlying settings. For example, simple models (such as linear regression or other parametric models) are often preferred for small sample sizes. For large sample sizes, more complex, nonlinear models can be used to capture heterogeneity in the treatment effect. Notwithstanding, the ability to handle nonlinear relationships and treatment effect heterogeneity is not unique to causal ML but can, in principle, also rely on classical statistical models that allow incorporating prespecified nonlinearities. Therefore, causal ML may have advantages when the underlying data-generating process is complex and when prior knowledge is limited.

classical statistics often assume knowledge about the parametric form of the association between patient characteristics and outcomes, such as linear dependencies. However, such knowledge is often not available or unrealistic, especially for high-dimensional datasets such as electronic health records, and this can easily lead to models that are misspecified. By contrast, causal ML typically allows for less rigid models, which helps in capturing complex disease dynamics as well as human pathophysiology and pharmacology. Still, there is a trade-off as causal ML typically requires larger sample sizes.

The fundamental problem of causal inference

Estimating treatment effects from data requires custom methods. This is because treatment effects for individual patients are not observable owing to the so-called fundamental problem of causal inference^{18,19}: that is, one can only observe the factual patient outcome under the given treatment, but one never observes the counterfactual patient outcome under a different, hypothetical treatment (Fig. 1b). Therefore, the estimation of treatment effects or other causal quantities that are based on such unobserved outcomes poses challenges that do not exist in traditional, predictive ML.

First, to obtain a causal quantity (such as response to treatment) that can be estimated, certain assumptions on the causal structure of the problem must be made. In particular, one often needs to assume that there is no unmeasured confounding; that is, there are no unobserved factors that drive both treatment decisions and subsequent patient outcomes. If unmeasured confounding is present, the estimated treatment effects may suffer from confounding bias and, as a result, can be incorrect²⁰. Additionally, to estimate treatment effects, one needs to account for the dependence structure between treatment, outcomes and patient characteristics by modeling the underlying causal relationships. This is because intervening on the treatment

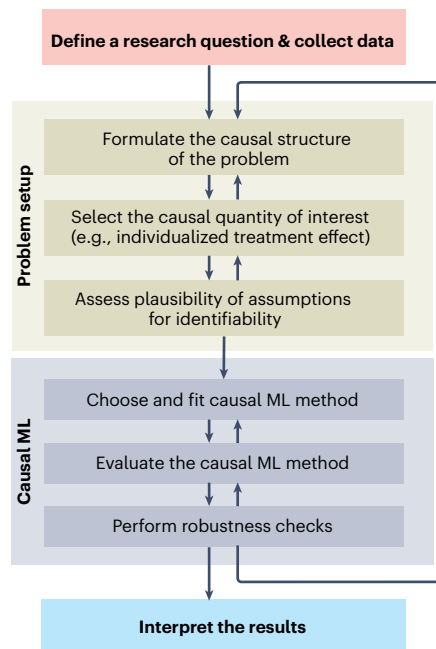


Fig. 2 | Workflow for causal ML in medicine. To predict treatment outcomes, assumptions on the causal structure of the problem must be made. This is relevant regardless of whether causal ML approaches or traditional statistical approaches are used. Subsequently, the causal quantity of interest can be predicted by causal ML.

variable could also affect other patient characteristics. As an example, consider a patient with a high body mass index whose doctor recommends quitting smoking, and for whom the diabetes risk should be predicted. Literature from traditional ML would suggest using both the body mass index and smoking behavior to predict the diabetes risk under a smoking versus no-smoking scenario; however, this approach would ignore that stopping smoking would also change a patient's body mass index. To address this issue, ML needs to be embedded in a causal framework.

The causal ML workflow

The process of predicting treatment outcomes with causal ML can be broken down into a few key steps (Fig. 2), which are discussed in the sections below. Following this workflow^{21,22} should help researchers to clearly define the research question and then guide their formulation of the problem structure, their choice of the causal quantity of interest, the causal ML method, the evaluation metric and the appropriate robustness checks to validate the reliability of the estimates.

Formulate the causal structure of the problem

To estimate the effectiveness of treatments, information about the following variables is necessary¹⁹: the treatment of interest, the observed patient outcome and patient characteristics (covariates) such as age, gender and the medical history. For example, in cancer care, one could use electronic patient records with information about the type of chemotherapy (the treatment), the size of a cancer tumor (the outcome), and the previous medical history (the covariates). In the standard setting¹⁹, the variables can influence each other as shown by the causal graph in Fig. 3a. To make causal quantities identifiable, we later need to assume knowledge about the causal graph.

Information about the above variables can come from either observational or experimental data. In observational data, such as clinical registries and electronic health records, the treatment assignment follows some typically unknown procedure, depending on the patient characteristics. For example, patients with a very severe illness are likely

to get a more aggressive form of treatment, implying that the patient characteristics differ across treatment groups. This contrasts with RCTs, where treatments are randomized and, as a result, the patient characteristics are similar across treatment groups. This is captured by the propensity score, which is the probability of receiving a treatment given the patient covariates¹⁴. In RCTs, the propensity score is known (for example, the propensity score is 50% in completely randomized trials with two treatment arms of equal size). By contrast, the propensity score in RWD is unknown, but it can be estimated to account for differences in the patient populations.

Select the causal quantity of interest

Causal quantities, such as the response to treatment, are commonly formalized based on the 'potential outcomes framework'¹⁵. The framework conceptualizes potential outcomes, which are the patient outcomes that would hypothetically be observed if a certain treatment was administered. Then, depending on the practical applications, different causal quantities can be of interest. These include treatment effects, which quantify the expected difference of two potential outcomes under different treatments. Common choices of treatment effects can be loosely grouped along two dimensions (Fig. 3b); the degree of effect heterogeneity and the treatment type. By choosing a specific treatment effect of interest, one defines the so-called estimand, that is, the causal quantity that should be predicted by the causal ML method.

Degree of effect heterogeneity. Traditionally, the average treatment effect (ATE) is widely used in clinical trials. The ATE measures effects at the level of the study population¹⁴. By comparing the average patient outcome for those receiving the treatment versus those who do not (control group), the ATE helps in understanding how effective a treatment is, on average, across a specific patient cohort²³. This is important, for example, when analyzing the comparative effectiveness of a new drug compared to the standard of care, or when assessing the overall effectiveness or safety of a new drug. However, the ATE cannot offer granular insights into whether patients with specific covariates may particularly benefit from a treatment, even though such heterogeneity in treatment effects can be of high interest in clinical practice (Fig. 1c). For a more granular view, one typically estimates the conditional average treatment effect (CATE), which is the effect of a treatment for a particular subgroup of patients defined by the covariates. Understanding the heterogeneity in treatment effects informs about subgroups where treatments are not effective or might even be harmful, which is relevant for individualizing treatment recommendations to specific patients.

Treatment type. Binary (discrete) treatments refer to a type of treatment variable that is dichotomous and thus has only two (or more) categories—for example, when answering questions of whether to treat or not to treat. By contrast, continuous treatments refer to a type of treatment variable that can take on a range of values rather than being limited to two (or a few) categories. Continuous treatment variables are commonly present in situations where the intensity, dosage or level of exposure to treatment can be flexibly chosen²⁴. For example, in radiation therapy, the dose of radiation is often chosen from a fairly wide spectrum that depends on the cancer type and other patient characteristics²⁵. For continuous treatments, the treatment effectiveness is often also summarized by dose–response curves.

Individual patient outcomes. Besides the above, some applications in medicine are also interested in predicting the individual patient outcomes. Predicting patient outcomes is different from treatment effects, as the former gives granular predictions of the potential outcomes under different treatments, while the latter estimates only comparative changes in outcomes but not the outcomes themselves. Therefore, treatment effects primarily tell the advantages of one treatment over another, while potential outcomes can support decision-making

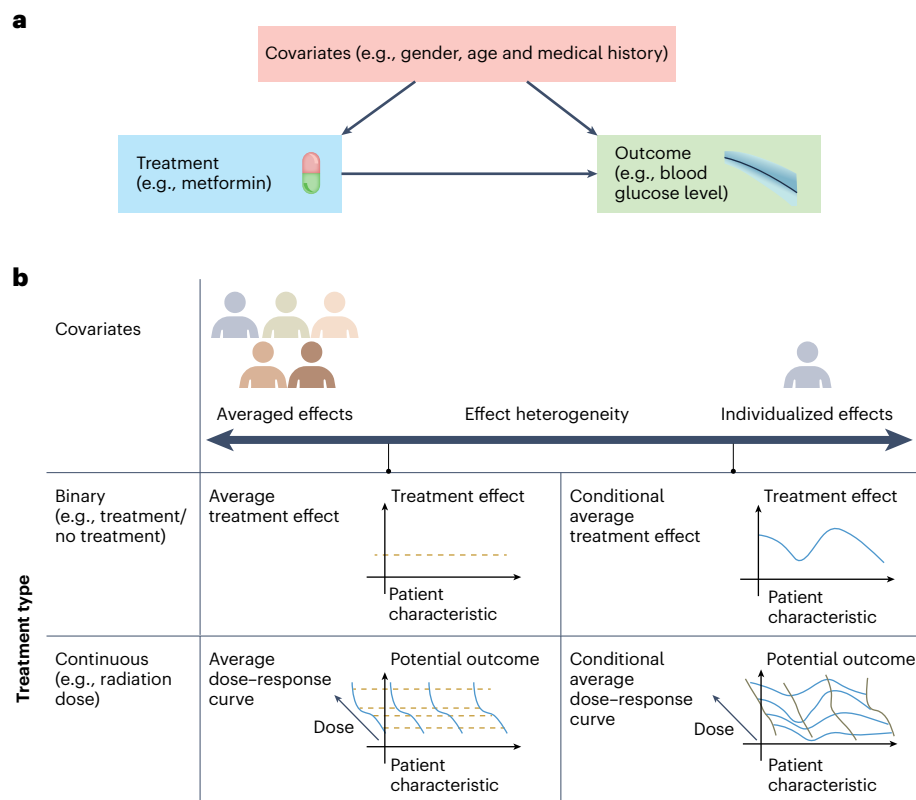


Fig. 3 | Formalizing tasks for causal ML. **a**, A causal graph must be assumed such as the example here. The arrows indicate the causal relationships between different variables. In the example, the assumption is made that patient outcomes are influenced by treatment and covariates. Note that the causal graph allows for possible unobserved variables (not confounders) that are correlated with treatment and confounders, or correlated with confounders and outcome.

b, The research question defines what causal quantity is of interest, that is, the estimand. The estimand can vary by the effect heterogeneity (average versus individualized) and treatment type (binary versus continuous). Depending on the causal graph and the causal quantity of interest, an appropriate causal ML method must be chosen.

in routine care by helping clinicians reason about what outcome to expect under different treatment options. This may be seen as a ‘risk under intervention’ estimand and requires a careful modeling strategy²⁶. For example, while the treatment effect may say that a drug can improve the 5-year mortality by five percentage points, the predicted outcomes could inform us that the mortality is 15% with treatment and 20% without. However, in practice, the estimation of ATE and CATE is often an easier task than predicting potential outcomes²⁷ and, hence, is preferred when it is sufficient for decision-making.

Assess the plausibility of assumptions for identifiability

The estimation of treatment effects involves counterfactual outcomes, which are not observable. Therefore, formal assumptions must be made about the data-generating process to ensure the identifiability of treatment effects from data¹⁹. Intuitively, identifiability is a theoretical concept that refers to whether causal quantities (such as treatment effects) can be uniquely inferred from data. Ensuring identifiability is a necessary step because, otherwise, it is impossible to estimate a treatment effect without bias, even with infinite data¹⁹.

RCTs ensure the identifiability of treatment effects through fully randomized treatment assignment. However, treatment assignment in RWD is not fully randomized and depends on covariates, so that formal assumptions must be made¹⁴. The exact set of assumptions depends on which type of treatment effect is chosen. For the treatment effects discussed above, in addition to having independent and identically distributed data, three ‘causal’ assumptions are standard^{14,28}. First, stable unit treatment value assumption (SUTVA) requires that the potential outcome coincides with the observed outcome for a given treatment and that the observed potential outcome on one patient

should be unaffected by the particular assignment of treatments to other patients. This assumption implies that there is no interference whereby treating one patient influences the outcomes for another patient in the study population (for example, due to spillover or peer effects). The SUTVA assumption also implies that there is hidden variation in the treatment effect across hospitals or populations. SUTVA is also known as consistency assumption together with non-interference. Second, positivity (also called overlap) requires a nonzero probability of receiving a treatment. Positivity implies that, for each possible combination of patient characteristics, we can observe both treated and untreated patients. And third, unconfoundedness (also called ignorability) states that, given observed covariates, the treatment assignment is independent of the potential outcomes. In particular, this is satisfied if the patient covariates include all possible confounders—in other words, variables that influence both the treatment and the outcome. For example, unconfoundedness may be violated if patients with certain sociodemographic characteristics (such as race or income level) tend to have better access to treatments²⁹, and where the reason is not captured in the data. In principle, unconfoundedness can be addressed by capturing all relevant factors driving treatment assignment in RWD³⁰, yet it is generally challenging to validate this in practice. If confounders are not observed or not modeled (or even not known), then the estimated treatment effect might be biased and thus incorrect²⁰.

Importantly, assumptions such as those above are required for consistently estimating treatment effects from data, regardless of whether a causal ML approach or a traditional statistical approach is used. A natural challenge comes from the fact that assessing the plausibility of the assumptions is often difficult. Later, we discuss potential

BOX 3

Model-agnostic methods for CATE estimation

There are different ways in which meta-learners can leverage the data in a supervised learning setting for CATE estimation.

Plug-in learners: One approach is to train a single ML model that predicts the patient outcome but where the treatment is added as a separate variable to the covariates (called S-learner⁵³). Another way is to train two separate ML models for each treatment (called T-learner⁵³). Here, one ML model is trained for predicting patient outcomes in the treatment group and one ML model for the control group. After having computed the ML model(s), one simply uses the estimated treated and control outcome to ‘plug them into’ the formula for computing the treatment effect.

Two-step learners: An alternative approach is to target the CATE, which can lead to faster convergence²⁷. However, because the difference between factual and counterfactual outcomes is never observed in data, so-called pseudo-outcomes are used as surrogates, which have the same expected value as the CATE. Prominent examples are the so-called DR-learner²⁷ and the so-called R-learner⁹⁸, which come with certain robustness guarantees^{61,98,99}.

The above meta-learners have different advantages and disadvantages. Unfortunately, there are no clear rules for choosing meta-learners but only high-level recommendations^{54,75,100}.

strategies to check the credibility of whether the assumptions hold. Notwithstanding, problem setups with alternative designs also exist. For example, some problem setups allow for relaxations of the SUTVA assumption (for example, by allowing for spillover effects)^{31,32}. There also exist alternatives to assuming unconfoundedness in specific settings, such as through the use of instrumental variables^{33,34}. Finally, there are problem setups that are not static but time-varying, so that a sequence of treatment decisions is made over time^{35–44}. Researchers are also developing ways to effectively combine both observational and experimental data^{45–47}.

Choose and fit the causal ML method

There are different causal ML methods, which vary based on which causal graph and which causal quantity of interest is addressed. For example, a large body of literature focuses on causal ML for ATEs^{23,48–52}. Here, a prominent method is based on so-called targeting to obtain an estimator that satisfies a semi-parametric efficient estimating equation. For CATE estimation with binary treatments, there are two broader categories of methods. On the one hand, so-called meta-learners⁵³ (Box 3) are model-agnostic methods for CATE estimation that can be used for treatment effect estimation in combination with an arbitrary ML model of choice (for example, a decision tree or a neural network⁵⁴). A key advantage of model-agnostic methods is that the underlying ML model can be chosen to flexibly handle clinical data sources such as electronic health records. On the other hand, model-specific methods make adjustments to existing ML models to address statistical challenges arising in treatment effect estimation and, therefore, to improve performance. Here, prominent examples that are particularly useful for clinical application are the causal tree⁵⁵ and the causal forest^{56,57}, which adapt the decision tree and random forest, respectively, for treatment effect estimation.

Even others adapt representation learning to leverage neural networks for treatment effect estimation^{58,59}. A different set of methods is needed for predicting the response to continuous treatment variables—for settings in which the intensity, dosage or level of exposure to a treatment can be flexibly chosen^{24,60–67}. This is because the number of treatment values is infinite and not every value is observed in the data—making treatment effect estimation particularly challenging in this context.

Existing causal ML methods often generate point estimates. This can be a serious limitation in medical applications⁶⁸, where uncertainty estimates such as standard errors or confidence intervals are crucial for reliable decision-making⁶⁹. However, there is also some progress. For example, for CATE estimation, the causal forest^{56,57} is a method that offers rigorous uncertainty estimates. In addition, several other strategies have been developed recently, such as Bayesian methods⁷⁰ and conformal prediction⁷¹, but still more research is needed.

Evaluate the causal ML method

Arguably, the best way to evaluate causal ML methods is to assess the accuracy in predicting patient outcomes from randomized data. While this does not allow assessment of treatment effects for individual patients, it still helps during model selection, so that models are favored with the best performance in terms of average or heterogeneous treatment effects. By contrast, benchmarking for the purpose of model selection is challenging, as both counterfactuals and ground-truth values of treatment effects are unknown^{72–74}. As a remedy, two strategies are common. A simple strategy is to compare methods from causal ML based only on the performance in predicting factual outcomes (whereby the performance in predicting counterfactual outcomes is ignored). This may give some insights into whether the underlying disease mechanisms in the data are captured. Yet it has a major limitation in that the key causal quantity of interest—that is, the treatment effect—is not evaluated. Another approach is to use pseudo-outcomes⁷⁵. Here, a pseudo-outcome is first estimated using a secondary, independent model to approximate the unknown counterfactual outcome, and then the pseudo-outcome is used to benchmark the estimated CATE. However, this approach depends on the performance of the secondary model for pseudo-outcomes and tends to favor certain methods⁷⁵. Overall, both strategies are merely heuristics and there is no ‘perfect’ solution.

Perform robustness checks

To validate the robustness of the treatment effect estimates against explicit violations of the different assumptions, so-called refutation methods are used⁷⁶. Common refutation methods include adding a random variable to check if the treatment effect estimates remain consistent (as such a variable should not affect the estimates), or replacing the actual treatment variable with a random variable to check if the estimated treatment effect goes to zero. Further, one could perform simulations where the outcome is replaced through semisynthetic data, to check if the treatment effect is correctly estimated under the new data-generating mechanism (for the simulated outcomes). Altogether, the choice of which refutation method to use for validating the causal ML methods highly depends on the specific problem setting and should be carefully chosen and implemented. Even when the refutation methods yield a positive result, this is no guarantee that the assumptions are satisfied. Nevertheless, robustness checks that are best practice in ML are still essential—for example, to mitigate the risk of bias⁷⁷—especially as the results in treatment effect estimation may heavily depend on both the data and the model choice.

Technical recommendations

To ensure the careful and reliable use of causal ML in clinical practice, we make several technical recommendations.

Checking the plausibility of assumptions

Assessing the plausibility of the underlying assumptions is crucial for the validity of treatment effect estimates, yet it is also challenging. For the consistency assumption, one should assert that the treatment of one patient does not affect the outcome of another based on domain knowledge. For the positivity assumption, one typically plots the propensity scores to check that they are not too small or too large; otherwise, there may not be enough support in the data for reliable inferences⁷⁸. Another strategy is to rely upon methods for uncertainty quantification as some treatments may be given rarely to certain patient cohorts, implying that there may be limited support in the data for making inferences in these patient cohorts and, therefore, a large uncertainty⁷⁹. If the positivity assumption is violated, one strategy is to exclude certain subgroups from the analysis as no reliable inferences for them can be made^{78,80}.

Validating the unconfoundedness assumption is especially challenging for RWD. The best way to avoid violations of the unconfoundedness assumption is to consult domain knowledge to ensure that all relevant factors behind treatment assignment are captured in RWD³⁰. An alternative is to adopt an instrumental variable approach^{33,34}; but appropriate instruments are often rare in medical applications and, again, the validity of instruments cannot be tested. If unobserved confounders cannot be ruled out, conducting a causal sensitivity analysis can be helpful to assess how robust the results are to potential unobserved confounding. Causal sensitivity analysis dates back to a study from 1959 showing that unobserved confounders cannot explain away the causal effect of smoking on cancer⁸¹. Causal sensitivity analysis computes bounds on the causal effect of interest under some restriction on the amount of confounding, thus implying that a treatment effect cannot be explained away. Restrictions on the amount of confounding are based on domain expertise, typically by making comparisons to known, important causes that act as baselines (for example, risk factors such as age). Recently, a series of causal ML methods have been proposed that provide sharp bounds^{82–86}. However, causal sensitivity analysis still requires that there is sufficient knowledge of human pathophysiology and pharmacology about important disease causes, which may not always be the case in observational studies²⁰.

Reporting

Findings should be interpreted and reported with great care. In particular, the assumptions, the rationale for the chosen causal ML method and the robustness checks should be clearly stated. If possible, the estimated treatment effects from RWD should be compared against those from RCTs. This can help in validating the reliability of the causal ML methods but may also reveal differences between clinical trials and routine care (for example, owing to different patient cohorts or different levels of adherence).

The reliability of the estimated treatment effects also depends on the quality and representativeness of the underlying data. Furthermore, analyses through causal ML involve multiple hypotheses testing and, therefore, are at risk of false positives. Similarly, owing to the retrospective nature of such analyses, another risk is selective reporting of positive results. To mitigate such risks, preregistered protocols for analysis are highly recommended^{87,88}. Finally, when causal ML is used together with RWD, the limitations of making causal conclusions should be openly acknowledged, and, if possible, RCTs should be considered for validation.

Clinical translation

By estimating treatment effects from medical data, causal ML offers substantial potential to personalize treatment strategies and improve patient health. Still, there is a long way to go. A key focus for future research must be on bridging the gap between ML research and direct benefits for patients in clinical practice.

Clinical use cases

Causal ML can help in generating new clinical evidence. For RCTs, causal ML may determine specific patient cohorts within the population that might respond positively (or negatively) to a particular treatment. For example, the treatment effect of antidepressant drugs compared to a placebo varies substantially and tends to increase with baseline severity of the depression⁸⁹. However, RCTs typically compare patient outcomes across two (or more) treatment arms, which would return the ATE at the population level, and the use of causal ML may help to define inclusion criteria for clinical trials or to identify predictive biomarkers (for example, certain genetic mutations in a tumor).

Furthermore, causal ML may offer flexible, data-driven methods to analyze treatment effect heterogeneity in RWD, including clinical registries and electronic health records. This is relevant as RCTs can be subject to limitations⁹⁰; for example, costs may be prohibitive or treatment randomization can be unethical for vulnerable populations (for example, pregnant women)⁹¹. RWD together with causal ML could allow the estimation of heterogeneous treatment effects for vulnerable groups, rare diseases, long-term outcomes and uncommon side effects that are often not sufficiently captured by traditional RCTs. For example, as randomizing hospitalizations is typically not possible, one study used causal ML to estimate the effect of hospitalizations on suicide risk from RWD⁹². Likewise, patient populations in RCTs are often not representative of the broader population⁹³, but one can account for this through causal ML⁹⁴ to better understand the post-approval efficacy of treatments. However, while the potential of RWD has been widely recognized^{90,95}, many methodological questions are still unanswered, and causal ML may thus help in translating data into clinical evidence.

Eventually, the choice of the specific estimand depends on the setting where causal ML is used. For regulatory bodies, it may be relevant to assess the overall net benefit for patients at large, for example, when comparing a new drug against the standard of care. This would require the estimation of the ATE. To ensure patient safety, regulatory bodies could also assess how the treatment effect varies across different subpopulations, which would involve the CATE. Likewise, the CATE may help to identify subpopulations that are particularly responsive to a treatment (for example, for hypothesis generation) or that would benefit from newly developed drugs, thereby contributing to an accelerated drug development. When causal ML is integrated into clinical decision support systems in routine care, clinical professionals may want to make personalized predictions of how a patient's health state changes under different treatment options. This would require methods for CATE estimation or even for predicting potential patient outcomes.

Challenges and future directions

Several challenges in the clinical translation of causal ML are at the technical level. First, both estimating heterogeneous treatment effects and predicting individual patient outcomes are naturally difficult. In practice, this often requires both strong predictors of treatment effects and large sample sizes. While the former depends on the human pathophysiology and pharmacology in the specific disease setup, the latter may improve over time with an increasing prevalence of electronic health records. Another challenge is that uncertainty quantification for many causal ML methods is lacking. However, uncertainty quantification is crucial for reliable decision-making and thus for building clinical evidence⁶⁹. For example, point estimates might indicate substantial effect heterogeneity, especially in settings with limited data, while in fact there may be little heterogeneity but simply large (aleatoric) uncertainty as the outcomes are difficult to predict. Hence, causal ML methods that only provide point estimates without conveying the appropriate uncertainty in the predictions may lead to potentially misleading or inappropriate conclusions. Finally, many causal ML methods are only implemented in specialized software libraries. Hence, comprehensive software tools are needed that improve reliability and ease of use, and

that account for practical needs in medicine (for example, rigorous uncertainty quantification).

The development of standardized protocols, ethical guidelines and regulatory frameworks for causal ML applications will be essential in ensuring safe and effective treatment decisions. For example, consensus-based, tailored checklists for reporting and quality will need to be developed. While there are checklists for traditional, predictive ML⁹⁶ and for generating real-world evidence^{88,97}, future research is needed that adapts such checklists to account for the needs of causal ML in medicine. Likewise, customized review processes will need to be developed, which define how evidence generated through causal ML methods must undergo regulatory review for approval.

So far, research in causal ML has primarily evaluated the performance of different methods through simulations (for example, refs. 35, 37,38,40,42,44). However, simulations involve (semi)synthetic datasets that do not fully capture the nuances of real-world disease dynamics. Hence, generating clinical insights through a cautious use of innovative causal ML methods can provide an important first step. This will help in understanding the strengths and limitations of causal ML in a medical context, especially in comparison to established clinical trial approaches. For this, settings where clear guidelines are missing could be appropriate, so that causal ML can provide input to augment the decision-making of clinical professionals. Causal ML for predicting treatment outcomes requires both methodological knowledge as well as domain knowledge of disease dynamics; therefore, cross-disciplinary collaboration between ML experts and clinicians is crucial for developing tools for clinical use. Eventually, tools based on causal ML may be integrated into routine care through clinical decision support systems. Such systems may directly predict individual patient outcomes for different treatment options and thereby support the decision-making of clinical professionals.

Conclusion

Causal ML offers the possibility to draw novel conclusions about the efficacy and safety of treatments and to personalize treatment strategies, thus improving patient health. However, in practice, several challenges arise, not least ensuring the reliability and robustness of these methods. Successful examples of causal ML in clinical use are still lacking, so proof-of-concept studies involving cautious use in clinical practice should be prioritized as an important first step.

References

- Kaddour, J., Lynch, A., Liu, Q., Kusner, M. J. & Silva, R. Causal machine learning: a survey and open problems. Preprint at arXiv <https://doi.org/10.48550/arXiv.2206.15475> (2022).
- Yoon, J., Jordon, J. & van der Schaar, M. GANITE: estimation of individualized treatment effects using generative adversarial nets. In *Proc. 6th International Conference on Learning Representations (ICLR)*, 2018.
- Evans, W. E. & Relling, M. V. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* **286**, 487–491 (1999).
- Esteva, A. et al. A guide to deep learning in healthcare. *Nat. Med.* **25**, 24–29 (2019).
- Kopitar, L., Kocbek, P., Cilar, L., Sheikh, A. & Stiglic, G. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Sci. Rep.* **10**, 11981 (2020).
- Alaa, A. M., Bolton, T., Di Angelantonio, E., Rudd, J. H. & van der Schaar, M. Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants. *PLoS ONE* **14**, e0213653 (2019).
- Cahn, A. et al. Prediction of progression from pre-diabetes to diabetes: development and validation of a machine learning model. *Diabetes/Metab. Res. Rev.* **36**, e3252 (2020).
- Zueger, T. et al. Machine learning for predicting the risk of transition from prediabetes to diabetes. *Diabetes Technol. Ther.* **24**, 842–847 (2022).
- Krittanawong, C. et al. Machine learning prediction in cardiovascular diseases: a metaanalysis. *Sci. Rep.* **10**, 16057 (2020).
- Xie, Y. et al. Comparative effectiveness of SGLT2 inhibitors, GLP-1 receptor agonists, DPP-4 inhibitors, and sulfonylureas on risk of major adverse cardiovascular events: Emulation of a randomised target trial using electronic health records. *Lancet Diabetes Endocrinol.* **11**, 644–656 (2023).
- Deng, Y. et al. Comparative effectiveness of second line glucose lowering drug treatments using real world data: emulation of a target trial. *BMJ Med.* **2**, e000419 (2023).
- Kalia, S. et al. Emulating a target trial using primary-care electronic health records: sodium glucose cotransporter 2 inhibitor medications and hemoglobin A1c. *Am. J. Epidemiol.* **192**, 782–789 (2023).
- Petito, L. C. et al. Estimates of overall survival in patients with cancer receiving different treatment regimens: emulating hypothetical target trials in the Surveillance, Epidemiology, and End Results (SEER)–Medicare linked database. *JAMA Netw. Open* **3**, e200452 (2020).
- Rubin, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66**, 688–701 (1974).
- Rubin, D. B. Causal inference using potential outcomes: design, modeling, decisions. *J. Am. Stat. Assoc.* **100**, 322–331 (2005).
- Robins, J. M. Correcting for non-compliance in randomized trials using structural nested mean models. *Commun. Stat.* **23**, 2379–2412 (1994).
- Robins, J. M. Robust estimation in sequentially ignorable missing data and causal inference models. In *1999 Proceedings of the American Statistical Association on Bayesian Statistical Science* 6–10 (2000).
- Holland, P. W. Statistics and causal inference. *J. Am. Stat. Assoc.* **81**, 945–960 (1986).
- Pearl, J. *Causality: Models, Reasoning, and Inference* (Cambridge University Press, 2009).
- Hemkens, L. G. et al. Interpretation of epidemiologic studies very often lacked adequate consideration of confounding. *J. Clin. Epidemiol.* **93**, 94–102 (2018).
- Dang, L. E. et al. A causal roadmap for generating high-quality real-world evidence. *J. Clin. Transl. Sci.* **7**, e212 (2023).
- Petersen, M. L. & van der Laan, M. J. Causal models and learning from data: integrating causal modeling and statistical estimation. *Epidemiology* **25**, 418–426 (2014).
- van der Laan, M. J. & Rubin, D. Targeted maximum likelihood learning. *Int. J. Biostatistics* **2**, 11 (2006).
- Hirano, K. & Imbens, G. W. in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family* (eds Gelman, A. & Meng, X.-L.) Ch. 7 (John Wiley & Sons, 2004).
- Specht, L. et al. Modern radiation therapy for Hodgkin lymphoma: field and dose guidelines from the international lymphoma radiation oncology group (ILROG). *Int. J. Radiat. Oncol. Biol. Phys.* **89**, 854–862 (2014).
- van Geloven, N. et al. Prediction meets causal inference: the role of treatment in clinical prediction models. *Eur. J. Epidemiol.* **35**, 619–630 (2020).
- Kennedy, E. H. Towards optimal doubly robust estimation of heterogeneous causal effects. *Electron. J. Stat.* **17**, 3008–3049 (2023).
- Imbens, G. W. & Rubin, D. B. *Causal Inference in Statistics, Social, and Biomedical Sciences* (Cambridge University Press, 2015).

29. Chen, J., Vargas-Bustamante, A., Mortensen, K. & Ortega, A. N. Racial and ethnic disparities in health care access and utilization under the Affordable Care Act. *Med. Care* **54**, 140–146 (2016).
30. Cinelli, C., Forney, A. & Pearl, J. A crash course in good and bad controls. *Sociol. Methods Res.* <https://doi.org/10.1177/00491241221099552> (2022).
31. Laffers, L. & Mellace, G. *Identification of the average treatment effect when SUTVA is violated. Department of Economics SDU. Discussion Papers on Business and Economics No. 3* (University of Southern Denmark, 2020).
32. Huber, M. & Steinmayr, A. A framework for separating individual-level treatment effects from spillover effects. *J. Bus. Econ. Stat.* **39**, 422–436 (2021).
33. Syrgkanis, V. et al. Machine learning estimation of heterogeneous treatment effects with instruments. In *Proc. 33rd International Conference on Neural Information Processing Systems* (eds Wallach, H. M. & Larochelle, H.) 15193–15202 (NeurIPS, 2019).
34. Frauen, D. & Feuerriegel, S. Estimating individual treatment effects under unobserved confounding using binary instruments. In *Proc. 11th International Conference on Learning Representations (ICLR, 2023)*.
35. Lim, B. Forecasting treatment responses over time using recurrent marginal structural networks. In *Proc. Advances in Neural Information Processing Systems 31* (eds Bengio, H. et al.) (NeurIPS, 2018).
36. Liu, R., Yin, C. & Zhang, P. Estimating individual treatment effects with time-varying confounders. In *Proc. IEEE International Conference on Data Mining (ICDM)* 382–391 (IEEE, 2020).
37. Li, R. et al. G-Net: a deep learning approach to G-computation for counterfactual outcome prediction under dynamic treatment regimes. In *Proc. Machine Learning for Health* (eds Roy, S. et al.) 282–299 (PMLR, 2021).
38. Bica, I., Alaa, A. M., Jordon, J. & van der Schaar, M. Estimating counterfactual treatment outcomes over time through adversarially balanced representations. In *Proc. 8th International Conference on Learning Representations* 11790–11817 (ICLR, 2020).
39. Liu, R., Hunold, K. M., Caterino, J. M. & Zhang, P. Estimating treatment effects for time-to-treatment antibiotic stewardship in sepsis. *Nat. Mach. Intell.* **5**, 421–431 (2023).
40. Melnychuk, V., Frauen, D. & Feuerriegel, S. Causal transformer for estimating counterfactual outcomes. In *Proc. 39th International Conference on Machine Learning* (eds Chaudhuri, K. et al.) 15293–15329 (PMLR, 2022).
41. Schulam, P. & Saria, S. Reliable decision support using counterfactual models. In *Proc. 31st International Conference on Neural Information Processing Systems* (eds von Luxburg, U. et al.) 1696–1706 (NeurIPS, 2017).
42. Vanderschueren, T., Curth, A., Verbeke, W. & van der Schaar, M. Accounting for informative sampling when learning to forecast treatment outcomes over time. In *Proc. 40th International Conference on Machine Learning* (eds Krause, A. et al.) 34855–34874 (PMLR, 2023).
43. Seedat, N., Imrie, F., Bellot, A., Qian, Z. & van der Schaar, M. Continuous-time modeling of counterfactual outcomes using neural controlled differential equations. In *Proc. 39th International Conference on Machine Learning* (eds Chaudhuri, K. et al.) 19497–19521 (PMLR, 2022).
44. Hess, K., Melnychuk, V., Frauen, D. & Feuerriegel, S. Bayesian neural controlled differential equations for treatment effect estimation. In *Proc. 12th International Conference on Learning Representations (ICLR, 2024)*.
45. Hatt, T., Berrevoets, J., Curth, A., Feuerriegel, S. & van der Schaar, M. Combining observational and randomized data for estimating heterogeneous treatment effects. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.2202.12891> (2022).
46. Colnet, B. et al. Causal inference methods for combining randomized trials and observational studies: a review. *Stat. Sci.* **39**, 165–191 (2024).
47. Kallus, N., Puli, A. M. & Shalit, U. Removing hidden confounding by experimental grounding. In *Proc. 32nd Conference on Neural Information Processing Systems* (eds Bengio, S. et al.) 10888–10897 (NeurIPS, 2018).
48. van der Laan, M. J., Polley, E. C. & Hubbard, A. E. Super learner. *Stat. Appl. Genet. Mol. Biol.* **6**, 25 (2007).
49. van der Laan, M. J. & Rose, S. *Targeted Learning: Causal Inference for Observational and Experimental Data* 1st edn (Springer, 2011).
50. Zheng, W. & van der Laan, M. J. In *Targeted Learning: Causal Inference for Observational and Experimental Data* 1st edn, 459–474 (Springer, 2011).
51. Díaz, I. & van der Laan, M. J. Targeted data adaptive estimation of the causal dose–response curve. *J. Causal Inference* **1**, 171–192 (2013).
52. Luedtke, A. R. & van der Laan, M. J. Super-learning of an optimal dynamic treatment rule. *Int. J. Biostatistics* **12**, 305–332 (2016).
53. Künzel, S. R., Sekhon, J. S., Bickel, P. J. & Yu, B. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc. Natl Acad. Sci. USA* **116**, 4156–4165 (2019).
54. Curth, A. & van der Schaar, M. Nonparametric estimation of heterogeneous treatment effects: From theory to learning algorithms. In *Proc. 24th International Conference on Artificial Intelligence and Statistics* (eds Banerjee, A. & Fukumizu, K.) 1810–1818 (PMLR, 2021).
55. Athey, S. & Imbens, G. Recursive partitioning for heterogeneous causal effects. *Proc. Natl Acad. Sci. USA* **113**, 7353–7360 (2016).
56. Wager, S. & Athey, S. Estimation and inference of heterogeneous treatment effects using random forests. *J. Am. Stat. Assoc.* **113**, 1228–1242 (2018).
57. Athey, S., Tibshirani, J. & Wager, S. Generalized random forests. *Ann. Stat.* **47**, 1148–1178 (2019).
58. Shalit, U., Johansson, F. D. & Sontag, D. Estimating individual treatment effect: generalization bounds and algorithms. In *Proc. 34th International Conference on Machine Learning* (eds Precup, D. & Teh, Y. W.) 3076–3085 (PMLR, 2017).
59. Shi, C., Blei, D. & Veitch, V. Adapting neural networks for the estimation of treatment effects. In *Proc. 33rd International Conference on Neural Information Processing Systems* (eds Wallach, H. M. et al.) 2496–2506 (NeurIPS, 2019).
60. Bach, P., Chernozhukov, V., Kurz, M. S. & Spindler, M. DoubleML: an object-oriented implementation of double machine learning in Python. *J. Mach. Learn. Res.* **23**, 2469–2474 (2022).
61. Foster, D. J. & Syrgkanis, V. Orthogonal statistical learning. *Ann. Stat.* **51**, 879–908 (2023).
62. Kennedy, E. H., Ma, Z., McHugh, M. D. & Small, D. S. Nonparametric methods for doubly robust estimation of continuous treatment effects. *J. R. Stat. Soc. Series B Stat. Methodol.* **79**, 1229–1245 (2017).
63. Nie, L., Ye, M., Liu, Q. & Nicolae, D. VCNet and functional targeted regularization for learning causal effects of continuous treatments. In *Proc. 9th International Conference on Learning Representations (ICLR, 2021)*.
64. Bica, I., Jordon, J. & van der Schaar, M. Estimating the effects of continuous-valued interventions using generative adversarial networks. In *Proc. 34th Annual Conference on Neural Information Processing Systems* (eds Larochelle, H. et al.) (NeurIPS, 2020).
65. Hill, J. L. Bayesian nonparametric modeling for causal inference. *J. Computational Graph. Stat.* **20**, 217–240 (2011).
66. Schwab, P., Linhardt, L., Bauer, S., Buhmann, J. M. & Karlen, W. Learning counterfactual representations for estimating individual dose-response curves. In *Proc. 34th AAAI Conference on Artificial Intelligence* 5612–5619 (AAAI, 2020).

67. Schweisthal, J., Frauen, D., Melnychuk, V. & Feuerriegel, S. Reliable off-policy learning for dosage combinations. In *Proc. 37th Annual Conference on Neural Information Processing Systems* (NeurIPS, 2023).
68. Melnychuk, V., Frauen, D. & Feuerriegel, S. Normalizing flows for interventional density estimation. In *Proc. 40th International Conference on Machine Learning* (eds Krause, A. et al.) 24361–24397 (PMLR, 2023).
69. Banerji, C. R., Chakraborti, T., Harbron, C. & MacArthur, B. D. Clinical AI tools must convey predictive uncertainty for each individual patient. *Nat. Med.* **29**, 2996–2998 (2023).
70. Alaa, A. M. & van der Schaar, M. Bayesian inference of individualized treatment effects using multi-task Gaussian processes. In *Proc. 31st Annual Conference on Neural Information Processing Systems* (eds von Luxburg, U. et al.) 3425–3433 (NeurIPS, 2017).
71. Alaa, A., Ahmad, Z. & van der Laan, M. Conformal meta-learners for predictive inference of individual treatment effects. In *Proc. 37th Annual Conference on Neural Information Processing Systems* (eds Oh, A. et al.) (NeurIPS, 2023).
72. Curth, A., Svensson, D., Weatherall, J. & van der Schaar, M. Really doing great at estimating CATE? A critical look at ML benchmarking practices in treatment effect estimation. In *Proc. 35th Conference on Neural Information Processing Systems Datasets and Benchmarks Track* (eds Vanschoren, J. & Yeung, S.-K.) (NeurIPS, 2021).
73. Boyer, C. B., Dahabreh, I. J. & Steingrimsson, J. A. Assessing model performance for counterfactual predictions. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.2308.13026> (2023).
74. Keogh, R. H. & van Geloven, N. Prediction under interventions: evaluation of counterfactual performance using longitudinal observational data. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.2304.10005> (2023).
75. Curth, A. & van der Schaar, M. In search of insights, not magic bullets: towards demystification of the model selection dilemma in heterogeneous treatment effect estimation. In *Proc. 40th International Conference on Machine Learning* (eds Krause, A. et al.) 6623–6642 (PMLR, 2023).
76. Sharma, A., Syrgkanis, V., Zhang, C. & Kiciman, E. DoWhy: addressing challenges in expressing and validating causal assumptions. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.2108.13518> (2021).
77. Vokinger, K. N., Feuerriegel, S. & Kesselheim, A. S. Mitigating bias in machine learning for medicine. *Commun. Med.* **1**, 25 (2021).
78. Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y. & van der Laan, M. J. Diagnosing and responding to violations in the positivity assumption. *Stat. Methods Med. Res.* **21**, 31–54 (2012).
79. Jesson, A., Mindermann, S., Shalit, U. & Gal, Y. Identifying causal-effect inference failure with uncertainty-aware models. In *Proc. 34th Conference on Neural Information Processing Systems* (eds Larochelle, H. et al.) 11637–11649 (NeurIPS, 2020).
80. Rudolph, K. E. et al. When effects cannot be estimated: redefining estimands to understand the effects of naloxone access laws. *Epidemiology* **33**, 689–698 (2022).
81. Cornfield, J. et al. Smoking and lung cancer: recent evidence and a discussion of some questions. *J. Natl Cancer Inst.* **22**, 173–203 (1959).
82. Frauen, D., Melnychuk, V. & Feuerriegel, S. Sharp bounds for generalized causal sensitivity analysis. In *Proc. 37th Annual Conference on Neural Information Processing Systems* (eds Oh, A. et al.) (NeurIPS, 2023).
83. Kallus, N., Mao, X. & Zhou, A. Interval estimation of individual-level causal effects under unobserved confounding. In *Proc. 22nd International Conference on Artificial Intelligence and Statistics* (eds Chaudhuri, K. & Sugiyama, M.) 2281–2290 (PMLR, 2019).
84. Jin, Y., Ren, Z. & Candès, E. J. Sensitivity analysis of individual treatment effects: a robust conformal inference approach. *Proc. Natl Acad. Sci. USA* **120**, e2214889120 (2023).
85. Dorn, J. & Guo, K. Sharp sensitivity analysis for inverse propensity weighting via quantile balancing. *J. Am. Stat. Assoc.* **118**, 2645–2657 (2023).
86. Oprescu, M. et al. B-learner: quasi-oracle bounds on heterogeneous causal effects under hidden confounding. In *Proc. 40th International Conference on Machine Learning* (eds Krause, A. et al.) 26599–26618 (PMLR, 2023).
87. Hernán, M. A. & Robins, J. M. Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* **183**, 758–764 (2016).
88. Xu, J. et al. Protocol for the development of a reporting guideline for causal and counterfactual prediction models in biomedicine. *BMJ Open* **12**, e059715 (2022).
89. Fournier, J. C. et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* **303**, 47–53 (2010).
90. Booth, C. M., Karim, S. & Mackillop, W. J. Real-world data: towards achieving the achievable in cancer care. *Nat. Rev. Clin. Oncol.* **16**, 312–325 (2019).
91. Chien, I. et al. Multi-disciplinary fairness considerations in machine learning for clinical trials. In *Proc. 2022 ACM Conference on Fairness, Accountability, and Transparency (FACCT '22)* 906–924 (ACM, 2022).
92. Ross, E. L. et al. Estimated average treatment effect of psychiatric hospitalization in patients with suicidal behaviors: a precision treatment analysis. *JAMA Psychiatry* **81**, 135–143 (2023).
93. Cole, S. R. & Stuart, E. A. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *Am. J. Epidemiol.* **172**, 107–115 (2010).
94. Hatt, T., Tschernutter, D. & Feuerriegel, S. Generalizing off-policy learning under sample selection bias. In *Proc. 38th Conference on Uncertainty in Artificial Intelligence* (eds Cussens, J. & Zhang, K.) 769–779 (PMLR, 2022).
95. Sherman, R. E. et al. Real-world evidence—what is it and what can it tell us. *N. Engl. J. Med.* **375**, 2293–2297 (2016).
96. Norgeot, B. et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat. Med.* **26**, 1320–1324 (2020).
97. Von Elm, E. et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Lancet* **370**, 1453–1457 (2007).
98. Nie, X. & Wager, S. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* **108**, 299–319 (2021).
99. Chernozhukov, V. et al. Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21**, C1–C68 (2018).
100. Morzywołek, P., Decruyenaere, J. & Vansteelandt, S. On a general class of orthogonal learners for the estimation of heterogeneous treatment effects. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.2303.12687> (2023).

Acknowledgements

S.F. acknowledges funding via Swiss National Science Foundation Grant 186932.

Author contributions

All authors contributed to conceptualization, manuscript writing and approval of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to Stefan Feuerriegel.

Peer review information *Nature Medicine* thanks Matthew Sperrin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Karen O’Leary, in collaboration with the *Nature Medicine* team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2024

Off-Policy Learning for Audience-Wide Content Promotions

Joel Persson

ETH Zurich

Stefan Feuerriegel

Munich Center for Machine Learning & LMU Munich

Cristina Kadar

Neue Zürcher Zeitung (NZZ)

Publishers and producers of digital content face the challenge of selecting which content to promote on their distribution channels (e.g., website front page, email newsletters, and social media pages), where a small set of items is shown to the entire audience, and experiments are often impractical due to the risks and opportunity costs of randomization. In this paper, we develop a framework for off-policy learning from historical data to address this problem. The objective is to, based on contextual information, decide which items to promote per channel over time so that the mean reward given capacity constraints is maximized. To enable off-policy learning in the absence of an experiment, we develop a model of the historical data-generating process and show that it implies non-parametric identification. We connect this to a causal machine learning procedure that is doubly robust against selection bias and reward misspecification and that discovers which contextual information informs the optimal decisions. In an empirical application at an international newspaper, our approach increases the revenue from promotions by USD 1.32–6.61 million compared to current practice. We isolate mechanisms for the improvement, and find that running a randomized experiment would have incurred a revenue loss from sub-optimal promotions that is several orders of magnitude greater than the revenue gain, demonstrating the economic importance of learning from historical data. Altogether, our work provides a robust, scalable, and cost-effective causal machine learning framework to evaluate and optimize audience-wide content promotion strategies.

Key words: content promotion; digital distribution channels; historical data; off-policy learning; causal machine learning

1. Introduction

Crucial for the success of content is how decision-makers distribute it to audiences.¹ This is especially important in online settings where some content receives vastly more views, clicks, and engagement than others (Berger and Milkman 2012, Upworthy 2012, Robertson et al. 2023). As such, publishers of online content are faced with selecting a few of their currently relevant items to promote to their audience on their distribution channels, such as the front page of their website, their social media pages, and email newsletters, so that key performance measures for the business (e.g., traffic, engagement, conversions) are maximized.

In this work, we study a common and practically important content selection problem: *audience-wide content promotions*. As a canonical example, consider a reputable newspaper like the New York Times. The decision problem of audience-wide content promotion involves selecting a few items from all currently relevant content items to populate on digital distribution channels, such as social media pages (where personalization by the page owner is infeasible) and homepages (where only a few items in banners and widgets are personalized but where the majority of content is uniformly curated for reputational and brand strategy reasons).

Audience-wide content promotions have a long history in practice, and differ from personalized content recommendations in both aims and solutions. Personalized recommendations, popularized by content aggregators and platforms like Netflix, YouTube, and Spotify, involve optimally distributing a vast pool of content from diverse producers to a broad population with *heterogeneous* preferences. In contrast, audience-wide promotions are primarily used by content publishers and producers (e.g., New York Times, Forbes, content creators) who own a smaller pool of content aimed at a clearly defined audience with largely *homogeneous* preferences.² Research by the New York Times has even shown that many of their users subscribe to the newspaper because they prefer unbiased curation over personalization (Rockwell 2019), thus speaking to the value of audience-wide promotions from a customer perspective. Previous research has thoroughly studied how to optimize personalized content recommendations (e.g., Agarwal et al. 2015, Ansari et al. 2018, Lada

¹ A study from 2020 found that US chief marketing officers considered distribution to be the single most important content marketing activity (EMarketer 2020a). This is supported by the fact that content marketing received 66.3% of all investments and, thus, more than other activities such as advertising, search engine optimization, and paid search (EMarketer 2020b).

² While content publishers and producers may also utilize personalized recommendations, they do so for different reasons than audience-wide promotions. Examples include the personalization of targeted ads on their channels (Kallus and Udell 2020), banners displaying next-read recommendations, opt-in subscription newsletters (Kadar 2022), and homepage widgets (Agarwal et al. 2015, Garcin et al. 2013).

et al. 2019, Zhang et al. 2019, Song et al. 2019, Rafieian 2022, Rafieian et al. 2023, Wang et al. 2023), but research on optimizing audience-wide content promotions remains scarce.

In this paper, we present a comprehensive off-policy learning framework for optimizing audience content promotions across digital distribution channels. The problem is to learn a decision policy of which content to promote on each channel such that the expected reward (in terms of traffic, engagement, or conversions) for yet unseen content is maximized. An intrinsic feature of audience-wide promotions is that the same promoted content is shown to the *entire audience*, and, therefore, experimentation may be undesired or infeasible due to opportunity costs and risks. As such, companies benefit from using historical data from past promotion decisions.

To enable the identification required for off-policy learning, we first develop a model of how historical data was generated under the current practice. Our central insight from the model is that the decision-maker faced an online learning problem requiring continual exploration over exploitation, thereby inducing identifying variation in promotion decisions. We use this to derive a formula for nonparametric identification from the historical data, connect it to the construction of an optimal promotion policy, and present a state-of-the-art causal machine learning procedure for estimation. We show that our empirical strategy is robust against lack of point identification of counterfactuals, selection bias in the historical data, and reward misspecification (Dudík et al. 2014, Athey and Wager 2021, Kennedy 2023), thereby accounting for common challenges for off-policy learning from observational data. We further synthesize advances in post-Lasso inference (Belloni and Chernozhukov 2013, Belloni et al. 2014, Zhao et al. 2021) with variable selection using multiple testing (Pelger and Zou 2022) to discover the drivers of the heterogeneity in promotion effects, and show that this enables reducing the information to base promotion decisions upon without sacrificing performance.

To evaluate our framework, we partnered with *Neue Zürcher Zeitung (NZZ)*, a leading news media company in Switzerland and Germany.³ The company’s goal is to decide which news stories their editors should promote on the front page for the Swiss and German markets to maximize a score of traffic, engagement, and subscriptions. We use a unique, large-scale dataset covering over 2000 news articles, 600 rounds of promotion decisions, and a rich set of covariates, including all information shown to the editors at the time of their decisions, all of their decisions, and all outcomes. Additionally, we construct covariates not available to the editors but which may serve as proxies for unmeasured heterogeneity and confounding, such as promotional histories, temporal

³ German front page: <https://www.nzz.ch/deutschland>. Swiss front page: <https://www.nzz.ch>

dynamics, and language sentiment of content extracted via a large language model (LLM) (Guhr et al. 2020, Guhr 2022).

In out-of-sample off-policy evaluations, our optimal policy increases revenue from ads, sales, and subscriptions by USD 1.32–6.61 million compared to current editorial practices, a statistically and economically significant improvement. We analyze the mechanisms behind this gain and find non-linear and high-dimensional heterogeneity regarding when, where, and which content our optimal policy promotes relative to the status quo. Drawing upon the logic of ablation studies in machine learning, we further use our framework to off-policy evaluate alternative counterfactual policies so as to isolate the gains and losses had alternative strategies been used. By comparing to a promotion policy that selects items uniformly at random, we find evidence of selection bias in the current practice, thus supporting our approach to account for this in learning from the historical data. The evaluation of the random policy also suggests that running a randomized experiment would have incurred a revenue loss from sub-optimal promotions of up to USD 14 million relative to current practice, outweighing the empirically best achieved revenue gain by a factor of 2.12–10.61. Thus, the standard approach of collecting completely randomized data for off-policy learning would have led to revenue loss that would not have recouped. In contrast, our approach based on historical data incurs no such revenue loss, and thus all of the revenue gain is a surplus.

Our contributions are as follows: Methodologically, we demonstrate that off-policy learning is feasible without data from a randomized experiment, provided a model of how the historical data to use has been generated, and that the identification and estimation is robust. We believe that this approach of connecting a model of the data-generating process (DGP) with identification approaches from the causal inference literature and robust estimation via machine learning can be valuable in other marketing contexts where randomized experiments are impractical but rich historical data is available. Previous research (Lada et al. 2019) has proposed the use of historical data to learn heterogeneous treatment effects for optimizing news recommender systems, but without showing formally how the data enables identification.

Second, building on the logic of ablation studies for evaluating machine learning algorithms, we show how off-policy evaluation can be used to causally compare the outcomes of marketing decision strategies and for attributing differences in their outcomes to their allocation mechanisms. Previous research in marketing (e.g., Smith et al. 2022, Hitsch et al. 2024) has studied how the outcomes of a fixed optimal policy change when evaluated by different machine learning models. We flip the logic and, instead, study how the outcomes change when the same machine learning models are used

to evaluate decision policies with different treatment assignment mechanisms. We show that this can be used to quantify the performance and revenue gains of marketing decision-making based on causal inference vs. outcome prediction, to discover simple heuristics that outperform expert decisions, and to estimate the economic opportunity cost of running a randomized experiment.

Third, we contribute to the growing body of empirical research using the incremental-effect approach to marketing decision-making, as seen in targeting and personalization studies (Bodapati 2008, Ascarza 2018, Lemmens and Gupta 2020, Hitsch et al. 2024). Here, we develop a comprehensive framework for the novel decision problem of *audience-wide* content promotions where, in contrast to targeting and personalization, randomized experiments are often undesirable due to costs, time, and risk. Finally, we offer managerial suggestions for improving content promotions in practice. Field experiments (Robertson et al. 2023) and conventional newsroom wisdom, such as “If it bleeds, it leads,” suggests that negative news generates more clicks. Our findings challenge this and thus call caution. Positive stories can be beneficial for optimizing long-term outcomes, which is crucial for digital content businesses (Yang et al. 2024).

The rest of the paper is structured as follows. In Sec. 2, we review previous research related to digital content optimization and off-policy learning. In Sec. 3, we present our framework for off-policy learning for optimizing content promotions. In Sec. 4, we apply our framework to the decision-making problem at our partner company. Finally, we discuss our contribution and managerial implications (Sec. 5).

2. Related Work

Our work contributes to previous research on optimizing digital content selection (e.g., Agarwal et al. 2009, Hauser et al. 2009, Kale et al. 2010, Li et al. 2010, Garcin et al. 2013, Urban et al. 2014, Schwartz et al. 2017). Due to the breadth of the literature, we focus on a few selected studies relevant to ours.

Learning how to optimize content selection can be done “online” or “offline”, also known as on-policy vs. off-policy learning. Online learning converges towards an optimal policy on-the-fly by making decisions and recording rewards as new observations arrive, thereby sequentially improving the decision policy based on the action-reward feedback. Offline learning, in contrast, infers an optimal policy without real interactions by leveraging historical data generated according to another policy (Sutton and Barto 2018, Ch.,5.4).

Methods for online learning have been studied extensively for the personalization of website design, content targeting, and news recommendations, typically using contextual bandit algorithms

and adaptive experiments (e.g., Agarwal et al. 2009, Hauser et al. 2009, Kale et al. 2010, Li et al. 2010, Garcin et al. 2013, Urban et al. 2014, Schwartz et al. 2017, Dimakopoulou et al. 2019). Because online learning necessitates balancing exploration with exploitation (Sutton and Barto 2018), they are feasible when the costs of exploration (or, more generally, randomization) are low. Personalization is such a setting, as then, sub-optimally selected arms are only shown to a few individual users. For audience-wide promotions, however, the same content is shown to everyone, making the costs of exploration prohibitive, for instance, in terms of reputational risk, degradation of user experience, or revenue loss.

Offline learning enables risk-free testing and optimization without experimenting on users, and is thus more suitable for audience-wide promotions. However, the feasibility of offline learning depends on the ability to identify counterfactuals from historical data. Previous research on content optimization from historical data has used structural econometric models (e.g., Johar et al. 2014, Besbes et al. 2016, Song et al. 2019, Zhang et al. 2019), which, although performing counterfactual analysis offline, may suffer from a lack of scalability and robustness. We contribute to this field by proposing a causal machine learning approach to off-policy learning, which benefit from being data-driven, robust to misspecification, and scalable to the types of data encountered in digital content settings (e.g., text data and high-dimensional contextual information).

Methods for policy learning, whether online or offline, can further differ in whether they select actions based on the prediction of outcomes from actions or the prediction of incremental effects (i. e., treatment effects) from actions. Outcome prediction has been the typical approach for research leveraging bandits and structural models (e.g., Agarwal et al. 2009, Hauser et al. 2009). However, recent research in marketing has shown that decision-making based on outcome prediction is generally sub-optimal, and that optimizing against incremental effects is more effective (e.g., Ascarza 2018, Lemmens and Gupta 2020, Hitsch et al. 2024). We contribute to this line of work by extending the incremental-effects approach to off-policy learning in a new empirical setting, namely audience-wide promotions.

Our work also relates to emerging marketing research on conditional average treatment effects (CATE) and off-policy evaluation using machine learning (e.g., Simester et al. 2020b, Liu 2022, Smith et al. 2022, Ellickson et al. 2023, Yoganarasimhan et al. 2023, Hitsch et al. 2024, Huang and Ascarza 2024, Yang et al. 2024). Some works use double machine learning (DML) to estimate CATE. While DML is doubly robust, its statistical inference is valid only if the estimated CATE includes the true CATE. We therefore leverage doubly robust learning (DRL) (Dudík et al.

2014, Foster and Syrgkanis 2019, Kennedy 2023), which instead of integrating machine learning with orthogonalization like DML, integrates machine learning with augmented inverse probability weighting (AIPW) (Robins et al. 1994, Robins 1999). DRL thus has the benefit over DML of allowing for statistical inference on treatment effect heterogeneity under weaker assumptions while being more robust and sample-efficient than standard inverse probability weighting (IPW). As such, we show that learning and optimizing audience-wide promotion strategies can be cast in an off-policy learning framework leveraging DRL.

3. Our Off-Policy Learning Framework

We now present our framework for off-policy learning. We first introduce a model of how the historical data was generated. We then formalize the objective of using that data for off-policy learning. Based on that, we derive the form of the optimal policy to learn and show how our model implies an identification approach permitting machine learning estimation. Finally, we provide a robust causal machine learning procedure for estimation, incorporating the discovery of effect heterogeneity and optimal policy variable selection.

3.1. A Model of the Data-Generating Process

In the absence of a randomized experiment, off-policy learning requires the following components: (1) a model of the DGP; (2) an identification formula for evaluating counterfactual policies given the DGP; and (3) a statistical estimator that maps the identification formula and the observed data to an policy value estimate. In the following, we present our approach to step (1). We consider a Bayesian myopic decision-maker (e. g., editor) who faced an online learning problem with Markovian dynamics. Later, we detail our approaches to (2) and (3).

Preliminaries: Let $t = 1, \dots, T$ denote time period, $i = 1, \dots, N_t$ content items, and $\mathcal{D}_n = \{(\mathbf{X}_{it}, \mathbf{A}_{it}, Y_{it}) : i \in \mathcal{I}_t, t = 1 \dots, T, n = \sum_t N_t\}$ the historical data, where each $(\mathbf{X}_{it}, \mathbf{A}_{it}, Y_{it})$ is an i.i.d. realization of a context, action, and reward drawn from an unknown stationary joint distribution P_π . In what follows, we present the components of this data-generating distribution.

Contexts: At the start of each time period $t = 1, \dots, T$, the editor receives a set \mathcal{I}_t of items $i = 1, \dots, N_t$, $N_t = |\mathcal{I}_t|$, that are relevant for the audience at the time. For each item $i \in \mathcal{I}_t$, the editor observes a context $\mathbf{X}_{it} = (X_{it1}, \dots, X_{itp})^T \in \mathcal{X} \subset \mathbb{R}^p$, which consists of p covariates (e.g., time information, time-invariant content characteristics, time-varying past performance indicators). Contexts represent observed heterogeneity across items and are i.i.d. draws from an unknown fixed distribution $P_{\mathbf{X}}$.

Actions: Conditional on a context \mathbf{X}_{it} , the editor decides whether to promote each item $i \in \mathcal{I}_t$ or not on each channel $m = 1, \dots, M$. Let $\mathbf{A}_{it} = (A_{it1}, \dots, A_{itM})^T \in \mathcal{A} = \{0, 1\}^M$ denote the vector of promotion decisions for an item across the M channels, where $A_{itm} = 1$ denotes that item i was promoted on channel $m = 1, \dots, M$ at time t , and $A_{itm} = 0$ denotes that it was not. For example, if there are $M = 2$ channels and an item i was promoted on both, then $\mathbf{A}_{it} = (1, 1)$. In the terminology of the literature, the promotion decisions correspond to treatment assignments, treatment arms, allocations, or actions, and so we will use these terms interchangeably where convenient. At most $C_{tm} < N_t$ items can be promoted on each channel $m = 1, \dots, M$ per time period t , as the pool of content \mathcal{I}_t is much larger than the capacity of a channel (e.g., in our empirical application, the newspaper has over 200 relevant articles at a time but can promote only a subset thereof). The constraint may vary by channel and over time to accommodate differential design and demand across surfaces (e.g., a desktop web page versus an email newsletter) and temporal dynamics (e.g., weekday versus weekend traffic). We denote the selected subset of items by $\mathcal{C}_{tm} = \{i \in \mathcal{I}_t : A_{itm} = 1, \sum_i A_{itm} = C_{tm}\}$.

Rewards: At the end of the time period, realizations of rewards $Y_{it} = Y(\mathbf{X}_{it}, \mathbf{A}_{it}) \in \mathcal{Y} \subset \mathbb{R}$ are recorded for each item $i \in \mathcal{I}_t$, which are i.i.d. draws from an unknown time-invariant conditional reward density $p(Y_{it} | \mathbf{x}_{it}, \mathbf{a}_{it})$.

Promotional value: Let $(\mathbf{x}_{it}, \mathbf{a}_{it}, y_{it})$ be an arbitrary context-action-reward realization. Given the described DGP, the likelihood of observing the realization is $p(\mathbf{x}_{it}, \mathbf{a}_{it}, y_{it}) = p(\mathbf{x}_{it})\pi(\mathbf{a}_{it} | \mathbf{x}_{it})p(y_{it} | \mathbf{x}_{it}, \mathbf{a}_{it})$. Let P_π denote the joint distribution function of the data associated with this likelihood function. The realized mean reward in the historical time period t is then given by

$$V_t(\pi) := \mathbb{E}_{(\mathbf{X}_t, \mathbf{A}_t, Y_t) \sim P_\pi} [Y(\mathbf{X}_t, \mathbf{A}_t)] = \int y(\mathbf{x}_t, \mathbf{a}_t) dP_\pi, \quad (1)$$

that is, the average item-reward given the promotion allocation decisions. We call this expectation $V_t(\pi)$ the time t *value* of the behavior policies π , as it quantifies how good the promotional decisions in that time period actually were in terms of the expected reward. The aim of the decision-maker was to maximize the long-term cumulative value,

$$V(\pi) := \sum_{t=1}^T V_t(\pi). \quad (2)$$

To facilitate the exposition of our identification strategy, we define probability densities as Radon-Nikodym derivatives with respect to a common dominating probability measure \mathbb{P} , i.e., $p_\pi = dP_{(\cdot)} / d\mathbb{P}$. We thereby assume without loss of generality that density functions are absolutely continuous with respect to \mathbb{P} .

3.2. Model Implications

We now discuss the implications of the model for how items were historically allocated to promotion and, hence, the usefulness of the historical data for off-policy learning. Because the historical promotion decisions were made on-the-fly as new items were sampled, online learning theory posits that the editors faced an exploration-exploitation trade-off (Sutton and Barto 2018).

- *Exploration:* This means that the editors promoted items more or less independent of their contexts or, similarly, that they (not) promoted items whose contexts have not yet been (had already been) subject to promotion. Thus, if an editor explored in a time period t , the allocation decisions were generated as

$$A_{itm}, \dots, A_{N_t, tm} \sim \text{Binom}(N_t, \mathbb{P}_\pi[A | \mathbf{X}]) \quad (3)$$

$$\text{s.t. } \sum_{i \in \mathcal{I}_t} A_{itm} \leq C_{tm}, \quad \text{for each channel } m = 1, \dots, M. \quad (4)$$

where, with slight abuse of notation, $\mathbb{P}_\pi[A | \mathbf{X}] \in (0, 1)$ denotes the unknown probability of a promotion decision given an item context, bounded away from zero and 1.

- *Exploitation:* Exploitation means that the editors promoted the items whose contexts were believed to map to greater potential rewards. The potential rewards are unknown at the time of the decisions, and so to circumvent the cold-start problem the editor (must at least implicitly have) relied on a subjective prior $\tilde{p}(Y(\mathbf{X}_{it}, \mathbf{a}_{it}))$, which maps a context-action pair per channel to a distribution over potential rewards. Thus, if an editor exploited, then the actions in $A_{1tm}, \dots, A_{N_t, tm}$ in the data for time period t are the solution to

$$\max_{\mathbf{a}_{1t}, \dots, \mathbf{a}_{N_t, t}} \int_{\mathcal{X}} \int_{\mathcal{Y}} y(\mathbf{X}_{it}, \mathbf{a}_{it}) d\tilde{P}_{Y(\mathbf{X}_{it}, \mathbf{a}_{it}) | \mathbf{X}_{it} = \mathbf{x}_{it}}(y_{it}) dP_{\mathbf{X}}(\mathbf{x}_{it}) \quad (5)$$

$$\text{s.t. } \sum_{i \in \mathcal{I}_t} A_{itm} \leq C_{tm}, \quad \text{for each channel } m = 1, \dots, M. \quad (6)$$

That is, under exploitation, the editor selected C_{tm} items to promote per channel $m = 1, \dots, M$ such that the prior belief over the value function was maximized, where the overall mean is obtained by marginalizing over contexts in the subjective rewards predictions.

From the lens of online learning theory, exploration vs. exploitation corresponds to strategic vs. greedy behavior, in that exploitation maximizes immediate rewards without enabling better future decision-making, whereas exploration acquires additional information to improve later decisions at the cost of foregoing immediate rewards. From a statistical perspective, exploration vs. exploitation corresponds to a random vs. deterministic treatment assignment. Thus, reliable identification of

counterfactuals as required by off-policy learning relies on that the editors did not always exploit. There are several reasons for why this holds true. On a general level, exploration is a necessary condition for online learning (Sutton and Barto 2018). Hence, to maximize the cumulative long-term promotional value, editors must necessarily have explored. On a more specific level, the structure of the problem that the editor faced inherently limits the ability for exploitation, thus forcing a degree of exploration. Reasons for this include:

- *Cold start*: Time periods may have contained new items with previously unseen contexts, because of which there is a need for exploration among editors to address the cold-start problem in the absence of an explicit prediction model.
- *Partial reward feedback*: Rewards are only recorded for the decisions that are made, i.e., $Y_{it} = \left(\prod_{m=1}^M \mathbb{1}\{A_{itm} = a_{itm}\} \right) Y(\mathbf{X}_{it}, a_{it1}, \dots, a_{itM})$. Hence, for each non-cold start item still relevant, the editor only knew one of the 2^M potential rewards.
- *Unknown context relevance*: The editors' priors on feature importance are weak, that is, they have imperfect knowledge of how or which context covariates map to greater rewards, and so exploration is required to acquire additional information.
- *Massive choice set*: The set of feasible allocation decisions in each time period is $\{\mathbf{a}_{it} \in \{0, 1\}^M : \sum_{i \in \mathcal{I}_t} a_{itm} \leq C_{tm}, m = 1, \dots, M\}$, for which possibly only one allocation maximizes the value. This set grows exponentially in the number of channels and multiplicatively in the number of items, thus demanding computational abilities for effective exploitation, far exceeding that of humans.
- *Heterogeneous editors*: In many application areas, the promotion decisions are made by multiple individuals with heterogeneous priors due to different expertise, preferences, and adherence to guidelines. Decision-makers may change over time and across items, causing the decisions in the item-hour data, even when conditioned on contexts, to vary exogenously due to the lack of information on who made them.⁴

Altogether, the above features of the historical DGP imply the data contain “as-good-as” random variation in promotion decisions that is exploitable for identification of counterfactuals. Technically, we have that $\mathbb{P}(A_{itm} = a \mid \mathbf{X}_{it} = \mathbf{x}) \neq \mathbb{P}(A_{jst} = a \mid \mathbf{X}_{js} = \mathbf{x})$ for any decision $a = 0, 1$ to items i, j in any two past time periods $t, s \leq T$ for any pair of channels $m, l \in [M]$. We thus next provide our objective for off-policy learning from the historical data. Later, we clarify how our model of the

⁴ A canonical example is the empirical setting of this paper: a newspaper where a team of heterogeneous editors makes decisions based on information from a dashboard. The same information is shown to everyone, but which editor promotes which item at which time is not a priori decided internally.

DGP implies a non-parametric identification formula permitting the use of machine learning for estimation with theoretical guarantees for unbiased and doubly robust off-policy learning. In doing so, we discuss the assumptions implied by the model and verify empirically that the propensity scores are varying over the probability range.

3.3. Off-Policy Learning Objective

To learn a new policy, we decompose the T time periods from the historical data into a training set \mathcal{T} for learning the policy and a test set \mathcal{V} for evaluating it, where $T = |\mathcal{T}| + |\mathcal{V}|$ and $s > t$ for all time periods $s \in \mathcal{V}$, $t \in \mathcal{T}$.

Let $d_m : \mathcal{X} \rightarrow \{0, 1\}$ be a stationary deterministic policy, meaning a mapping from the context space to the space of promotion decisions for a channel $m = 1, \dots, M$, where $d_m(\mathbf{X}_{it}) = 1$ means that the policy promotes an item with context \mathbf{X}_{it} on channel m , and $d_m(\mathbf{X}_{it}) = 0$ that it does not. We consider separate policies for each channel, since the same items are not necessarily best to promote on all channels. The value of a policy d_m at time t is given by

$$V_t(d_m) = \mathbb{E}_{P_{d_m}}[Y(\mathbf{X}_{it}, A_{tm}, \mathbf{A}_t^{-m})] = \int Y(\mathbf{x}_t, A_{tm} = d_m(\mathbf{x}_t), \mathbf{A}_t^{-m}) dP_{d_m}, \quad (7)$$

where we decompose the decisions \mathbf{A}_t to all M channels into those for channel $m = 1, \dots, M$, i.e., A_{tm} , and those for the other $M - 1$ channels, i.e., \mathbf{A}_t^{-m} . This allows us to consider how rewards change as a function of decisions for only a single channel while the decisions to the other channels are held constant. The distribution P_{d_m} refers to the joint distribution of the data if the decisions to channel m would be made according to a policy d_m , but, for the other channels, the decisions would be made as in the historical data.

The above estimand $V_t(d_m)$ allows for comparing the value of a policy per time period, as we can expect there to be temporal heterogeneity in expected outcomes. To evaluate a policy overall, we use the long-run average value

$$V(d_m) = \frac{1}{|\mathcal{V}|} \sum_{t \in \mathcal{V}} V_t(d_m), \quad (8)$$

which averages out the temporal heterogeneity in value over the $|\mathcal{V}|$ time periods in the test set \mathcal{V} .

To this end, the objective is to, for each channel $m = 1, \dots, M$, use historical data from the training set \mathcal{T} to learn an optimal policy for the test set \mathcal{V} , meaning a counterfactual policy that solves the following constrained optimization problem:

$$d_m^* := \arg \max_{d_m \in \mathcal{D}} V(d_m) \quad (9a)$$

$$\text{s.t. } \sum_{i \in \mathcal{I}_t} d_m(\mathbf{X}_{it}) \leq C_{tm}, \quad \text{for all } t \in |\mathcal{V}|. \quad (9b)$$

Here, $\mathcal{D} \subseteq \{\mathcal{X} \rightarrow \mathcal{A}\}$ is a finite set of policies represented by a chosen function class of machine learning models. Thus, the optimal policy attains the maximum value on the test set among those admissible with the machine learning model used, that is, $V(d_m^*) \geq V(d_m)$ for any policy d_m in \mathcal{D} .

3.4. Optimal Policy

We now show that the optimal policy is a simple threshold rule on the CATE. The CATE measures heterogeneity around the average treatment effect as a function of covariates. In our setting, we have multiple binary decision variables (i.e., whether to promote an item or not on each of the channels) that jointly affect the outcome. As such, we note that the CATE of promotion on a single channel $m = 1, \dots, M$ can be isolated from the joint CATE given promotion on all M channels as

$$\tau_m(\mathbf{x}) := \mathbb{E}[Y(\mathbf{X}_{it}, A_{itm} = 1, \mathbf{A}_{it}^{-m}) - Y(\mathbf{X}_{it}, A_{itm} = 0, \mathbf{A}_{it}^{-m}) \mid \mathbf{X}_{it} = \mathbf{x}]. \quad (10)$$

This CATE estimand is the counterfactual difference in expected outcomes had an item been promoted on channel m or not given its context and the decisions to the other channels. The following theorem establishes the optimal promotion policy for a channel given the capacity constraint.

THEOREM 1. *Let $\tau_m(\mathbf{x})$ denote the CATE to channel $m = 1, \dots, M$ given context $\mathbf{X} = \mathbf{x}$ according to Eq. (10) and let $\tau_{tm}^{(C_{tm})}$ be the C_{tm} -th largest CATE among those among the $|\mathcal{I}_t|$ items relevant at time t . The optimal policy is*

$$d_m^*(\mathbf{x}) = \mathbb{1} \left\{ \tau_m(\mathbf{x}) \geq \tau_{tm}^{(C_{tm})} \right\} = \begin{cases} 1, & \text{if } \tau_m(\mathbf{x}) \geq \tau_{tm}^{(C_{tm})}, \\ 0, & \text{else.} \end{cases} \quad (11)$$

where potential ties in CATE across items within a time period are broken at random.

A proof is provided in Appendix B.2. Theorem 1 states that, if only, say, $C_{tm} = 30$ items out of 200 can be promoted on a channel at a time, then the optimal decision is to promote the 30 items with the largest increment in the reward over their baseline if they would be promoted. Promoting those items will always maximize the mean reward irrespective of its form, as those are the items with greatest contribution to the overall mean (i.e., the objective function to maximize) across all items. It follows immediately that, for each time period $t \in \mathcal{V}$ and channel $m = 1, \dots, M$, the optimal set is given by $\mathcal{C}_{tm}^* = \{i \in \mathcal{I}_t : d_m^*(\mathbf{x}_{it}) = 1, \sum_{i \in \mathcal{I}_t} d_m^*(\mathbf{x}_{it}) = C_{tm}\}$.

3.5. Identification

We now discuss how our model of the DGP allows for identification. As researchers, we cannot observe the editor's prior, which contextual information they used to make the decisions, or whether

they explored or exploited. To avoid making unverifiable assumptions about unobservables, we take a non-parametric approach and view the editor’s decisions as arising from some unknown *behavior policies* $\pi = (\pi_1, \dots, \pi_M)$, where π_m is the behavior policy (or expert policy if the decisions are made by human professionals, as is the case here) for channel $m = 1 \dots, M$. Formally, a behavior policy is a mapping from the space of contexts \mathcal{X} (and priors) to the space of probability distribution over actions. Intuitively, it encodes the heuristics, directives, or preferences that guided the editor’s decisions. With slight abuse of notation, we will let $\pi(\mathbf{a}_{it} \mid \mathbf{x}_{it}) = \mathbb{P}[\mathbf{A}_{it} = \mathbf{a}_{it} \mid \mathbf{X}_{it} = \mathbf{x}_{it}]$ and $\pi_m(\mathbf{a}_{itm} \mid \mathbf{x}_{it}) = \mathbb{P}[\mathbf{A}_{itm} = \mathbf{a}_{itm} \mid \mathbf{X}_{it} = \mathbf{x}_{it}]$ denote the conditional probability of the observed actions in the data for all M channels and for channel $m = 1 \dots, M$, given a realized context \mathbf{x}_{it} , respectively, and let $\mathbf{A}_{it} \stackrel{\text{i.i.d.}}{\sim} \pi$, $\mathbf{A}_{itm} \stackrel{\text{i.i.d.}}{\sim} \pi_m$ for all items $i \in \mathcal{I}_t$ per channel $m = 1, \dots, M$. We connect our model to the following standard assumptions in the potential outcomes framework for causal inference (Imbens and Rubin 2015, Hernán and Robins 2020).

ASSUMPTION 1. (1) Consistency: $Y_{it} = Y(\mathbf{X}_{it}, \mathbf{A}_{it})$ for all $\mathbf{A}_{it} \in \{0, 1\}^M$ and $t = 1, \dots, T$. (2) Sequential unconfoundedness: $Y(\mathbf{X}_{it}, a_{itm}, \mathbf{A}_{it}^{-m}) \perp\!\!\!\perp \mathbf{A}_{it} \mid \mathbf{X}_{it}$ for all $a_{itm} \in \{0, 1\}$, $t = 1, \dots, T$, and $m = 1, \dots, M$. (3) Decision preserving confounding: if $\mathbb{E}[\hat{\tau}_m(\mathbf{x})] \geq \mathbb{E}[\hat{\tau}_{tm}^{(C_{tm})}]$ then $\tau_m(\mathbf{x}) \geq \tau_{tm}^{(C_{tm})}$ for all $\mathbf{x} \in \mathcal{X}$, $t = 1, \dots, T$, and $m = 1, \dots, M$. (4) Positivity: $0 < \pi_m(a_{itm} \mid \mathbf{X}_{it}) < 1$ for all $m = 1, \dots, M$, $\mathbf{X}_{it} \in \mathcal{X}$, $\mathbf{A}_{it} \in \mathcal{A}$, and $t \leq T$, such that $p(\mathbf{X}_{it}) > 0$.

Assumption 1.1 states that realized rewards correspond to the potential outcomes under the treatments assigned. This assumption holds by the partial reward feedback in the DGP. The assumption further implies that rewards only depend on their own treatment assignments, thereby ruling out interference. This is reasonable in our setting since, the audience generally engages with an item because precisely that item appeals to them, not because some other item did. Assumption 1.2 states that, given the context, the potential reward is independent of the current assignment.⁵ This assumption is standard in the literature on causal machine learning and is appears quite reasonable in our setting, as we control for all information shown to editors at the time of their decisions in addition to all history not observable to them. Nonetheless, we emphasize that for identifying our optimal policy, the assumption can be relaxed to that of Assumption 1.3, which allows for unmeasured confounding as long as the items with the C_{tm} highest estimated CATEs for channel m at time t have the corresponding C_{tm} highest true CATEs, in expectation. This is

⁵ The assumption is also known as sequential exogeneity, sequential ignorability, and selection on observables. In sequentially randomized trials such as adaptive A/B tests, Assumption 1.3 holds by design. In observational studies, treatments are not completely randomized, and, hence, sequential unconfoundedness is an assumption.

a sufficient condition, as all that is required to make the optimal decision for an item is that its CATE lies on the correct side of the true decision boundary. In that sense, our framework does *not* require accurate point identification or point estimation of the CATE. This provides a theoretical robustness guarantee when we learn from historical data, which may be subject to selection bias in promotion decisions that affect point predictions but not necessarily relative rank orders or CATEs. In a broader sense, the assumption is similar to but not as strong as that of Lada et al. (2019), who formalize sufficient conditions on the structure of unmeasured confounding due to omitted variable bias that imply that the CATE estimates preserve their true ranks. Lada et al. (2019) show that this “rank-preserving confounding” assumption (and thus our weaker variant) is useful when experimentation is infeasible but one has access to large amounts of observational panel data containing high-dimensional features with unknown features importances, all of which applies to our model and empirical setting. Finally, Assumption 1.4 states that all items in the data had a non-zero probability of being promoted or not. In Sec. 3.2, we have already shown why this is a consequence of our model of the DGP; see our discussion therein. In practice, the true propensities are not known, but can be non-parametrically estimated. This is not a weakness, as recent research in causal inference (Su et al. 2023) has shown that, in the absence of a randomized experiment, using estimated propensity scores for estimating treatment effects (and thus our optimal policy) is superior to using true propensity scores, had they been known. Later, we provide descriptive evidence supporting the assumption for our empirical application (Sec. 4.3).

To identify the value $V(d_m)$ of a counterfactual policy d_m from the historical data, we must be able to write $V(d_m)$ as an expectation with respect to the joint data distribution P_π implied by the behavior policies $\pi = (\pi_1, \dots, \pi_M)$. The following proposition establishes this non-parametrically.

PROPOSITION 1 (Identification formula). *Under our model and its implied Assumption 1, we have*

$$V_t(d_m) = \int \left[\underbrace{\mu(\mathbf{X}_t, d_m(\mathbf{X}_t), \mathbf{A}_t^{-m})}_{\text{conditional mean reward}} + \underbrace{\frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{itm} | \mathbf{X}_t)} \times (y_t - \mu(\mathbf{X}_t, A_{tm}, \mathbf{A}_{it}^{-m}))}_{\text{debiasing term}} \right] dP_\pi, \quad (12)$$

and thus $V(d_m) = T^{-1} \sum_{t=1}^T V_t(d_m)$, where $\mathbb{1}\{\cdot\}$ is the indicator function and $\mu(\mathbf{X}_t, d_m(\mathbf{X}_t), \mathbf{A}_{it}^{-m}) := \mathbb{E}[Y_t | \mathbf{X}_t, d_m(\mathbf{X}_t), \mathbf{A}_{it}^{-m}] = \int y_t dP_{Y_t | \mathbf{X}_t, d_m(\mathbf{X}_t), \mathbf{A}_{it}^{-m}}(y_t)$ is the reward function.

The proof uses a change of measure, deriving the likelihood ratio of the data under the behavior policies π and counterfactual policy d_m , and then re-writing the expression; see Appendix B.1.

Eq. (12) is an augmented inverse probability weighting (AIPW) formula (Robins et al. 1994, Robins 1999) for causal inference from observational data, tailored to our setting. Intuitively, the

formula shows that when the counterfactual and behavior policies differ, the counterfactual policy value equals the conditional mean reward at its decision, with contexts integrated out, analogous to any approach that only models outcomes. When the policies agree, the conditional mean reward is augmented with a debiasing term – which weights the error of the conditional reward by the inverse propensity score – to mitigate reward misspecification and selection bias. This debiasing term penalizes cases where the decision of the counterfactual policy to evaluate coincides with the decision of the behavior policy but (i) the conditional mean reward function is inaccurate, (ii) the propensity for decision in the data was low, or (iii) both (i) and (ii) apply.

Estimators based on AIPW offer two advantages. First, compared to standard inverse probability weighting (IPW), AIPW estimators are sample efficient (Robins and Rotnitzky 1995). This efficiency arises because AIPW estimators model both the rewards and the actions, whereas IPW only models actions. Thus, AIPW estimators utilize all of the observations in the data, whereas IPW estimators only utilize observations for which the behavior policy and the counterfactual policy agree. This benefit is important in our setting, as only a small subset of items can be promoted per time period, and, therefore, the likelihood of policy agreement is low. Second, AIPW estimators are doubly robust, which means they provide unbiased estimation of counterfactual policy values even if one of the conditional mean reward model or the propensity score model is incorrect (Robins et al. 1994), thereby adding to their strength for historical data.

Double robust methods may not be a significant benefit for learning from observational data if one uses parametric models, as the likelihood that the DGP conform to the imposed parametric structure is low, and thereby the doubly robustness condition is not achieved. By establishing non-parametric identification, we allow for non-parametric estimation of the rewards and propensity scores using machine learning, thereby increasing the likelihood of achieving double robustness. In addition, if one does not establish non-parametric identification from the DGP, then there are no guarantees that machine-learned policy estimates actually correspond to the counterfactual estimand of interest. It is also important to note that off-policy learning using only a machine learning model of rewards without incorporating propensity scores requires one to make the so-called realizability assumption, which states that true reward function is contained in the function class of the machine learning model used (Foster et al. 2018). Being doubly robust, AIPW does not require this assumption. Motivated by these considerations, we next present our causal machine learning procedure based on AIPW.

3.6. Estimation and Evaluation

With our identification strategy and optimal policy construction, all that remains is the estimation and evaluation of the optimal policy, provided next.

3.6.1. Overview. Our approach is based on splitting the historical data into a training set and a test set at a specific time point. The initial set of observations forming the training set is used to learn the models used for off-policy evaluation whereas the test set forms the sample with which we actually carry out the off-policy evaluation. Via the train-test split, we evaluate the extent to which our methods generalize to yet unseen content, which is how learned decision-policies are deployed in practice and how off-policy evaluation is used in the literature (Murphy 2005).

Based on this train-test approach, our identification formula in Proposition 1 leads to the following sample plug-in estimator for evaluation on the test data:

$$\widehat{V}(\widehat{d}_m^*) = \frac{1}{|\mathcal{V}|} \sum_{t \in \mathcal{V}} \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} \widehat{\Gamma}_{it}(\widehat{d}_m^*). \quad (13)$$

where

$$\widehat{\Gamma}_{it}(\widehat{d}_m^*) = \widehat{\mu}(\mathbf{X}_{it}, \widehat{d}_m^*(\widehat{\mathbf{X}}_{it}), \mathbf{A}_{it}^{-m}) + \frac{\mathbb{1}\{A_{itm} = \widehat{d}_m^*(\widehat{\mathbf{X}}_{it})\}}{\widehat{\pi}_m(A_{itm} | \mathbf{X}_{it})} (Y_{it} - \widehat{\mu}(\mathbf{X}_{it}, A_{itm}, \mathbf{A}_{it}^{-m})) \quad (14)$$

and $\widehat{d}_m^*(\mathbf{x}_{it})$ equals one if $\widehat{\tau}_m(\mathbf{x}_{it}) \geq \widehat{\tau}_{tm}^{(C_{tm})}$, and zero otherwise. Here, the empirical mean is taken over the so-called doubly robust scores $\widehat{\Gamma}_{it}(\widehat{d}_m^*)$ (Athey and Wager 2021) from the optimal policy estimate \widehat{d}_m^* on the hold-out test set \mathcal{V} . The doubly robust score is essentially a context-action-specific sample analog of the AIPW formula, that involves conditional reward predictions of a model $\widehat{\mu}$, conditional propensity score predictions from a channel-specific model \widehat{p}_{im} , and predictions of the optimal policy decisions $\widehat{d}_m^*(\widehat{\mathbf{X}}_{it})$ given predictions of the CATE. Given that these models are estimated according to best practices (explained later), the following two propositions hold.

PROPOSITION 2. *Under Assumption 1.3, the estimated optimal policy decision is unbiased of the true optimal policy decision, i.e.,*

$$\mathbb{E}[\widehat{d}_m^*(\mathbf{X}_{it})] = d_m^*(\mathbf{X}_{it}). \quad (15)$$

THEOREM 2. *Under Assumption 1, the above plug-in estimator is doubly robust, i.e.,*

$$\mathbb{E}[\widehat{V}(\widehat{d}_m^*)] = V(d_m^*). \quad (16)$$

We omit the proof to Proposition 2 as we already discussed its sufficient conditions in Sec. 3.5. The proof to Theorem 2 is available in Appendix B.3. Proposition 2 states that an estimated optimal policy will, on average, make the same decisions as the oracle. Theorem 2 implies that our

estimate thereof on the hold-out test set recovers its true performance on average, even if either the reward model or the propensity score model is incorrect. Thus, our evaluation is doubly robust. Together, these results provide theoretical guarantees that permits non-parametric estimation and counterfactual evaluation via machine learning. Overall, the procedure builds on the theory of doubly robust learning (see, e. g., Dudík et al. 2014, Foster and Syrgkanis 2019, Kennedy 2023, for details), which adapts the AIPW estimator of average treatment effects using parametric models to the estimation of so-called doubly robust scores of the potential outcomes using machine learning. In what follows, we describe the estimation of the machine learning models to plug into the doubly robust score and, hence, the counterfactual value estimator.

3.6.2. Nuisance models. To estimate the performance of any counterfactual policy from the historical data, we must undo the selection by the behavior policies and account for how rewards depend on contexts and actions. Proposition 1 shows that the value function depends on a propensity score model π_m per channel and a reward model μ . In the causal inference literature, these are referred to as nuisance models, as they are not of interest in themselves but are simply used to plug-in predictions in the value function estimator per Eq. (13). As such, for channel $m = 1, \dots, M$, we seek to learn nuisance models

$$\hat{\pi}_m(a_{itm} | \mathbf{x}_{it}) = \hat{\mathbb{P}}[A_{itm} = a_{itm} | \mathbf{X}_{it} = x_{it}], \quad (17)$$

$$\hat{\mu}(\mathbf{x}_{it}, a_{itm}, \mathbf{a}_{it}^{-m}) = \hat{\mathbb{E}}[Y_{it} | \mathbf{X}_{it} = x_{it}, A_{itm} = a_{itm}, \mathbf{A}_{it}^{-m} = \mathbf{a}_{it}^{-m}]. \quad (18)$$

Thus, we need one propensity score model per channel but only one reward model, where the latter follows from that $\mathbf{A}_{it} = (A_{itm}, \mathbf{A}_{it}^{-m})$ for any m , and, so, we can use the same reward model for different channels by simply interchanging which decision variable across the channels acts as the treatment and which act as controls. The only assumptions required for this approach to yield unbiased estimates of the doubly robust scores are that the reward error $\varepsilon_{it} = Y_{it} - \mu(\mathbf{X}_{it}, A_{itm}, \mathbf{A}_{it}^{-m})$ is conditionally exogenous within time periods (i.e., $\mathbb{E}[\varepsilon_{it} | \mathbf{X}_{it}, A_{it}] = 0$) and the multiple treatments have separably additive effects without interactions.⁶ Both hold by the structure imposed by our model and Assumption 1.2.

As the nuisance models are non-parametric, any machine learning model can, in principle, be used as function approximators. We estimate the nuisance models as gradient-boosted trees (Friedman

⁶ Intuitively, the latter means that the CATE of promoting an item on one channel cannot depend on whether it is promoted on another channel. This also greatly simplifies the model, since, without interactions, one does not need to predict the possible outcomes from every possible combination of promotion decisions across the channels. This is also reasonable from a practical point of view since promotions for different geographical markets are managed by different teams (and visitors are directly assigned to one front page based on their geographic location).

2001). Gradient-boosted trees have many practical advantages in that they are scalable, accommodate non-linearities in functional form, and are robust to overfitting (Friedman 2001). We use a log-loss for the propensity score model and a mean square error loss for the reward model. We trim the estimated propensity scores at the recommended threshold (i.e., $\exp(-5)$) (Battocchi et al. 2019) to address possible violations to positivity (Assumption 1.4) and to reduce the variance in the doubly robust scores caused by division with low propensities. We also experimented with random forests, which had slightly inferior predictive performance but led to qualitatively similar conclusions. Implementation details (e.g., hyperparameter tuning) are provided in Appendix D.

3.6.3. CATE model. We also use a model for the CATE to estimate the optimal policy decision for plugging into $\hat{d}^*(\mathbf{X}_{it})$ in Eq. (13). To fit a model of the CATE, we first need nuisance estimates of the CATEs per context to use as the dependent variable. A key idea here is that we can again leverage doubly robust scores, but instead of computing the doubly robust score for a policy (which we do not yet have but wish to learn), we compute the doubly robust scores of promotion and no promotion and take their difference as a doubly robust estimate of the CATE per item-time observation. Recall that we define the CATE as a function of intervening on one of the multiple treatment variables while holding the others fixed at their values in the data. We thus compute the doubly robust scores per item $i \in \mathcal{I}_t$ and time period t independently per channel $m = 1, \dots, M$ as

$$\hat{\Gamma}_{it}(a_m) := \hat{\mu}(\mathbf{X}_{it}, a_m, \mathbf{A}_{it}^{-m}) + \frac{\mathbb{1}\{A_{itm} = a_m\}}{\hat{\pi}_m(A_{itm} | \mathbf{X}_{it})} (Y_{it} - \hat{\mu}(\mathbf{X}_{it}, A_{itm}, \mathbf{A}_{it}^{-m})) \quad \text{for } a_m = 1, 0, \quad (19)$$

where $A_{itm}, \mathbf{A}_{it}^{-m}$ are the observed (i.e., factual) actions in the data. We then take the difference in these scores of for $a_m = 1$ and $a_m = 0$ as nuisance CATE estimates, i.e.,

$$\hat{\tau}_{itm} = \hat{\Gamma}_{it}(1) - \hat{\Gamma}_{it}(0). \quad (20)$$

Finally, we project these CATE estimates per channel on a non-parametric regression function $h_m: \mathcal{X} \mapsto \hat{\tau}_{itm}$ conditioned on contexts to obtain a model of how the heterogeneity in CATE for a channel depends on contexts:

$$h_m(\mathbf{x}) = \hat{\mathbb{E}}[\hat{\tau}_{itm} | \mathbf{X}_{it} = \mathbf{x}]. \quad (21)$$

Modeling the CATE serves three purposes. First, it allows for more robust, lower variance CATE predictions than just using doubly robust scores from the nuisance models. The CATE estimates obtained by the differences in doubly robust scores will generally be highly complex and noisy, as they depend on a nonlinear AIPW formula involving predictions from two different machine

learning models. Projecting the CATE estimates onto another machine learning model h_m reduces their complexity and variance via additional regularization and smoothing and is therefore more robust out-of-sample (Chernozhukov et al. 2018b, Kennedy 2023, Hitsch et al. 2024). The variance reduction is especially important in our framework, as we learn from historical data via AIPW and, as such, the nuisance CATE estimates may have excess variance due to division by low treatment propensities. The CATE modeling approach addresses this. Second, it facilitates deployment at scale. Having fitted a model h_m for the CATE, we only need to evaluate a single lower-dimensional model h_m at prediction time for obtaining the promotion decisions of the optimal policy. In contrast, using the nuisance models requires evaluating both the propensity score model and the reward model, which will generally also require more features as they are not regularizing contexts for the CATE, but for actions and rewards. Third, fitting the model h_m to predict CATE estimates with regularization on the contexts allows us to discover which covariates from the contexts explain heterogeneity in the incremental effects of promotion per channel. The doubly robust scores from the nuisance models do not permit such discovery, as they depend on contexts in a highly non-linear way in the AIPW formula involving two different models.

We integrate the modeling of the CATE into our AIPW approach via orthogonal random forests, which are a variant of causal forests (Wager and Athey 2018) and generalized random forests (Athey et al. 2019) that achieves a lower estimation error by using a doubly robust moment equation (Oprescu et al. 2019). We fit the orthogonal random forests per channel by first fitting a random forest (Breiman 2001) to predict the nuisance model CATE estimates $\hat{\tau}_{itm}$ from the context variables \mathbf{X}_{it} . We construct these random forests from 500 non-parametric regression trees using the sub-sampled “honest” procedure of Athey and Imbens (2016) and Wager and Athey (2018), where we use half of the observations for creating the tree structures and the other half for estimation. The orthogonal random forest is then estimated by solving the following local moment equation to minimize mean-square error of the CATE per context covariate profile:

$$\hat{\tau}_m^0(\mathbf{x}) = \arg \min_{\tau_m^0} \frac{1}{|\mathcal{T}_2|} \sum_{t \in \mathcal{T}} \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} \Phi_m(\mathbf{x}, \mathbf{X}_{it}) \left(\hat{\tau}_m(\mathbf{X}_{it}) - \tau_m^0 \right)^2, \quad (22)$$

where $\mathcal{T}_2 = \bigcup_{l=1}^L \mathcal{T}_2^{(l)}$ is the union of held-out data from the cross-validation folds part of our sample-split cross-fitting procedure (see Sec. 3.6.4 for a description and Appendix D.4 for details), $\hat{\tau}_m(\mathbf{X}_{it})$ is the nuisance CATE estimate given by the difference in doubly robust scores in Eq. (20), and $\Phi_m(\mathbf{x}, \mathbf{X}_{it})$ is a data-driven kernel function (or, similarity metric) that calculates how often context \mathbf{x} fall into the same leaf of the regression trees in the random forest as another item of content i

in time period t with context \mathbf{X}_{it} . The resulting CATE estimates $\hat{\tau}_m^0(\mathbf{x})$ are asymptotically Gaussian distributed and thereby allow for constructing asymptotically valid inference via bootstrap (Oprescu et al. 2019). We run the whole estimation procedure on the training data separately for the CATE functions of either front page. See Appendix D.2 for hyperparameter tuning.

3.6.4. Sample-split cross-fitting. The only requirement for good statistical properties (e.g., asymptotic normality and semi-parametric efficiency) of the final estimates from the CATE model is that the machine learning models are estimated with an appropriate sample-split cross-fitting procedure (van der Laan and Rose 2011, Chernozhukov et al. 2018a, Kennedy 2023). At a high level, sample-split cross-fitting means that different splits of the training data are used to estimate the nuisance models vs. constructing the doubly robust scores using their predictions, and the estimation of the CATE model. We combine this with cross-validation in the following manner:

We first split the training set of the historical data at random into two equally sized partitions. We then take one of the partitions and split it further into random cross-validation folds. On each such fold, we then fit a reward model and a propensity score model. We then take the other partition of the training set and also split it at random into as many cross-validation folds. For each content item in each such fold, we construct the doubly robust score using the predictions of a reward model and a propensity score model from different folds in other partitions of the training set. This is the cross-validated estimation of the CATE using the nuisance models. We repeat this last step until we have predicted the CATE for all items that we randomly allocated to the second partition of the training set. We then use these observations to fit the CATE model. The procedure is formalized in Appendix D.4.

Intuitively, the above procedure makes the final CATE predictions independent of the predictions from the nuisance models and, thereby, leads to reductions in bias and variance compared to non-sample split estimation, and allows for valid statistical inference although they are data-driven. The hyperparameter tuning of the models is optimized as part of sample-split cross-fitting; see Appendix D for details.

4. Empirical Application

4.1. Overview

To evaluate our framework, we partnered with *Neue Zürcher Zeitung* (NZZ), a leading newspaper in Switzerland and Germany. A key task at *Neue Zürcher Zeitung* is to optimize the audience-wide promotions for two of their distribution channels: the Swiss front page of the website and the German front page of the website. The front pages target different audiences from different

countries, and, hence, separate promotion policies should be learned for each front page. Our aim is thus to learn which content to promote per time period and front page so that the promotional value in terms of revenue and a proprietary score of performance (explained later) is maximized.

The decision-making process for content promotion at our partner company is as follows: Every hour, editors use an internal dashboard displaying data for news articles published on the website in the last 72 hours. The data includes performance metrics per content from previous periods (such as total clicks and average engagement time) and whether an article has been previously promoted. Based on this contextual information, editors select which articles to promote on the Swiss and German front pages. Typically, they aim to fill about 30 slots per hour, though up to approximately 60 articles can potentially be featured on either front page simultaneously. Screenshots of the front pages are available in Appendix A.

4.2. Data

We obtained access to internal data from our partner company spanning six weeks at the end of 2021. Our dataset includes all articles published during this period, along with the information presented to editors at the time of their promotion decisions, the decisions themselves, and the resulting rewards. Additionally, we have metadata on all published content. This dataset forms an unbalanced panel, where each article is observed over 72 consecutive time periods starting from its publication and ending when it is no longer considered relevant by the newspaper. We focus our analysis on observations from 6:00 a.m. to 11:00 p.m., as promotions are not conducted during nighttime hours. In collaboration with our partner company, we identified a comprehensive set of covariates, which include those shown in the internal dashboard in addition to ones that were not shown:

- *Time information:* We construct dummy variables for the hour of the day and day of the week. For each item, we construct variables with the total duration (in hours) since publication and total duration (in hours) on either front page.
- *Content characteristics:* We compute the content length (i. e., number of characters) of each item and, using the metadata, construct dummy variables for the following: section (e. g., Sport, Business & Finance, International), type (e. g., Opinion, Comment, Interview), format (e. g., regular, visual, opinion), etc. There are 16 sections, 11 types, and 4 different formats. Inspired by earlier research in marketing and management science (Archak et al. 2011, Berger and Milkman 2012, Berger et al. 2020), we estimate the sentiment of the title shown on the website, of the title shown in search engines (i. e., the ranked list obtained after a Google

search), and of the lead text summarizing the story on the front page. For this, we use a large language model (LLM) based on the BERT model (Kenton and Toutanova 2019), which was additionally pre-trained on 1.834 million German texts (Guhr et al. 2020, Guhr 2022). We use the LLM to classify the three types of texts of each content as positive, neutral, or negative and then construct a dummy variable with neutral being the reference category.

- *Past performance indicators:* We include time-varying variables of historical performance per item related to traffic, engagement, and conversions, both for the previous period alone and from the time period of publication up until the end of the previous time period. Traffic indicators include the number of clicks, whereas engagement indicators include average reading time, average scroll depth, and the recirculation rate (i.e., the proportion of views that were followed by at least one more view, thus accounting for continued engagement within user session up until the start of the current time period). The conversion indicators attribute the view of an item to a conversion, and include cumulative last-touch registrations (i.e., the count of times an article was the last view of a user before they registered an account) and subscription path (i.e., the proportion of times the article was among the 10 last articles a user viewed before buying a subscription).
- *Past promotion decisions:* To control for a delay and decay in the effect of past promotion decisions, we construct the first lag of the promotion decisions and their cumulative count since publication. To account for the two different front pages, we construct two binary variables indicating if an item was promoted on the Swiss front page and on the German front page. To control for other decisions for items, we construct variables indicating if an item was distributed via an email newsletter, a push notification, or on Twitter in the previous time period, as well as the cumulative count of these since publication.

All covariates are measured per time period *prior* to the promotion decisions. We thus preserve the chronological order in the decision process and ensure that there is no look-ahead bias or post-treatment bias. We one-hot-encode categorical variables. Altogether, our final dataset is an unbalanced panel with over 116,000 item-hour observations spanning $T = 603$ time periods, $N = 2189$ items, and 80 covariates. Summary statistics are provided in Appendix C.

The reward of interest is the proprietary performance score (for which a larger value is better) that our partner company computes as a weighted function of traffic, engagement, and conversions.⁷

⁷ The partner company computes the score by first scaling the performance indicators to a common unit for aggregation and then aggregates them to a scalar measure via a weighing function, where the weights are set according to their relative importance to the business (which we are not allowed to disclose to ensure that information on the different revenue streams remains confidential). The resulting performance score is non-negative unbounded with a

The performance score may be interpreted as the company’s utility function or as an item-hour surrogate of the company’s aggregate long-term revenue from ads and subscriptions, which are only observed infrequently at a company level, not per item and time period. The way in which the performance score is computed can further be motivated theoretically by the scalarization technique commonly used in multi-objective optimization.

4.3. Descriptives

We perform a descriptive analysis of the historical data. The analysis provides model-free evidence in support of our model of the historical DGP and documents sources of heterogeneity in the data that can be leveraged for off-policy learning.

Fig. 1 shows kernel density estimates of the empirical propensity score distributions for both front pages, where the empirical propensity score is the share of times an item was promoted in the data. Two findings are worth pointing out. First, the propensities cover about the full probability range. Most mass is concentrated at a propensity of 0.15–0.20 or less, which is explained by that the average time period contained 191 articles of which 32 were promoted on either front page (the propensity scores have mean=0.17 and SD=0.04). Second, the densities of the propensity scores overlap. The implication is that the data is about equally informative for learning CATEs and counterfactual policies for both front pages. Overall, Fig. 1 shows that the probability of promotion varied across items and time, thus providing empirical evidence to the implications of our model of the DGP (cf. Sec. 3.2 and Assumption 1.4 of overlap).

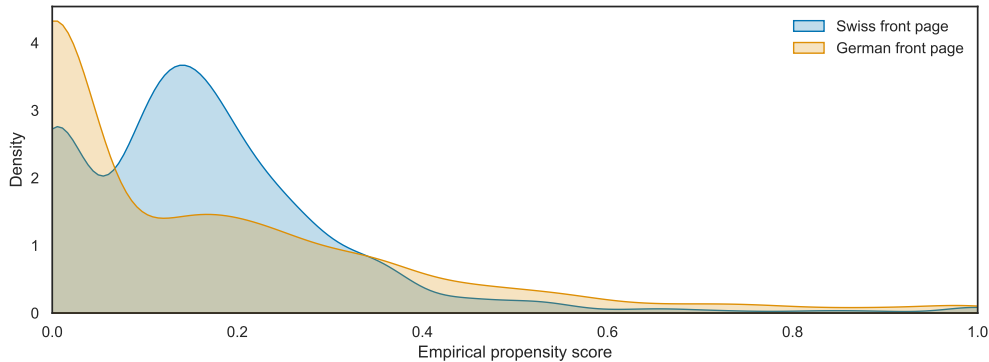


Figure 1 Distribution of empirical propensity scores (smoothed with kernel density estimation) of promotion.

Next, we document sources of heterogeneity explaining the variation in promotion propensities. The promotion propensities vary by the section, type, and format of the items (Fig. 2). For example, mean of 1.40 (for the Swiss front page) and 1.29 (for the German front page) and a standard deviation of 1.30 and 0.90, respectively.

for the Swiss front page, items of type “Explained” (i.e., stories that explained news events such as COVID-19) had a high propensity to be promoted, whereas for the German front page items by the editor-in-chief had a particularly high propensity of promotion. Moreover, items with Swiss domestic news were rarely promoted on the German front page, as Swiss news is typically not relevant to the German audience. We also find heterogeneity in promotion propensity with respect to the content sentiment we derive from the LLM (Fig. 3). For the lead text of content, promotion propensity was historically lower for more negative sentiment, whereas, for the title of content on the front page or in search engines, the propensity was largely uniform across negative, neutral, and positive sentiment. The editors were not shown the sentiment of content in the dashboard but they could have inferred it. We therefore include the sentiment in our nuisance and CATE models to control for unobserved heterogeneity in promotion effectiveness and thus for confounding.

On the one hand, the variation in promotion propensities across contexts (w.r.t. Section, Type, Format, or sentiment) indicates selection bias. On the other hand, the standard deviations around the conditional mean propensities suggests an exploration of rewards conditional on contexts. Together, this lends empirical support to our approach of leveraging the variability in actions for identification while accounting for selection bias via machine-learned AIPW estimation.

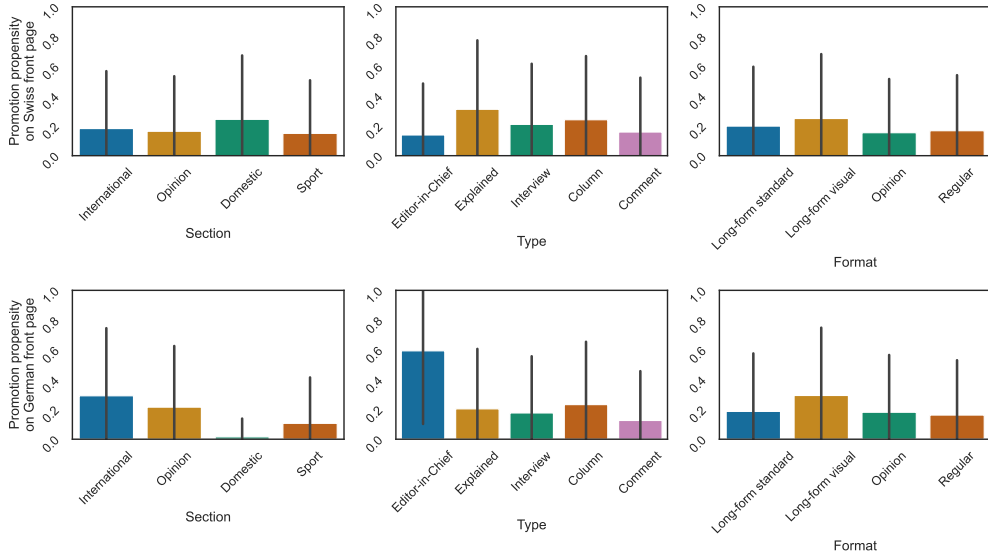


Figure 2 Historical mean promotion propensity by section, type, and format of an item. Error bars are standard deviations across 1000 bootstrap runs. Top: Swiss front page. Bottom: German front page.

Finally, we turn to heterogeneity in the performance score given the promotion decisions. We find that items that were promoted had a substantially larger performance score, and that this

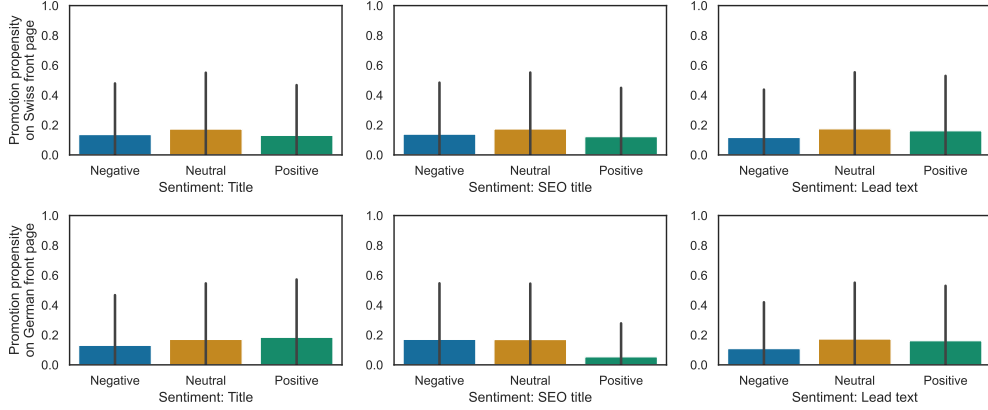


Figure 3 Historical mean promotion propensity by sentiment. Error bars are standard deviations across 1000 bootstrap runs. Top: Swiss front page. Bottom: German front page.

statistically significant but non-causal lift varies over hours of a day, days within weeks (Fig. 4), across content characteristics (Fig. 5), and across the two front pages. This suggests that contexts moderate both temporal and cross-sectional heterogeneity in promotion effectiveness along several covariates, and that the strength of the moderation per covariate further differs between the two front pages.

To summarize, the data appears to support our model of the DGP, our identification and estimation procedures, and our approach of learning separate policies per channel to account for differential promotion effectiveness and differential machine learning selection of controls and moderators. Having established that the data is informative and reliable to use for off-policy learning, we next apply our framework.

4.4. Ablation Study of Performance and Revenue Implications

4.4.1. Alternative promotion policies for comparison. We apply our framework to evaluate the optimal policy and alternative promotion policies common in practice and the literature. We off-policy evaluate these alternative policies using the same AIPW estimator given the same fitted nuisance models as for the optimal policy. Hence, any difference in promotional value between the policies we evaluate can only be attributed to their treatment assignment mechanisms. Overall, our off-policy evaluation approach to compare different promotional allocation mechanisms is analogous to the use of ablation studies for attributing the causal impact of different components in a machine learning algorithm to their predictive performance. Here, however, we use it to causally analyse how different promotion decision policies leads to different expected outcomes. The alternative policies we evaluate are:

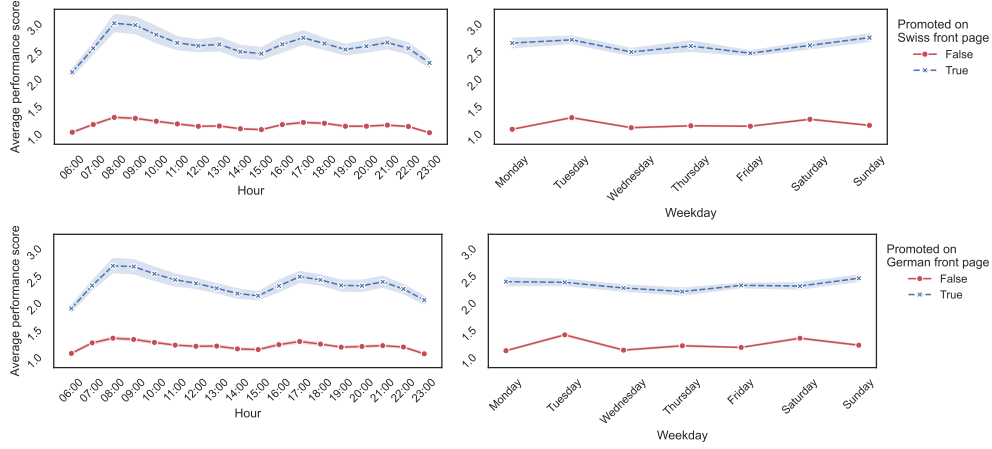


Figure 4 Average performance score by hour and day of the week across promoted and non-promoted items.

Shaded regions are 95% confidence intervals from 1000 bootstrap runs. Top: Swiss front page. Bottom: German front page.

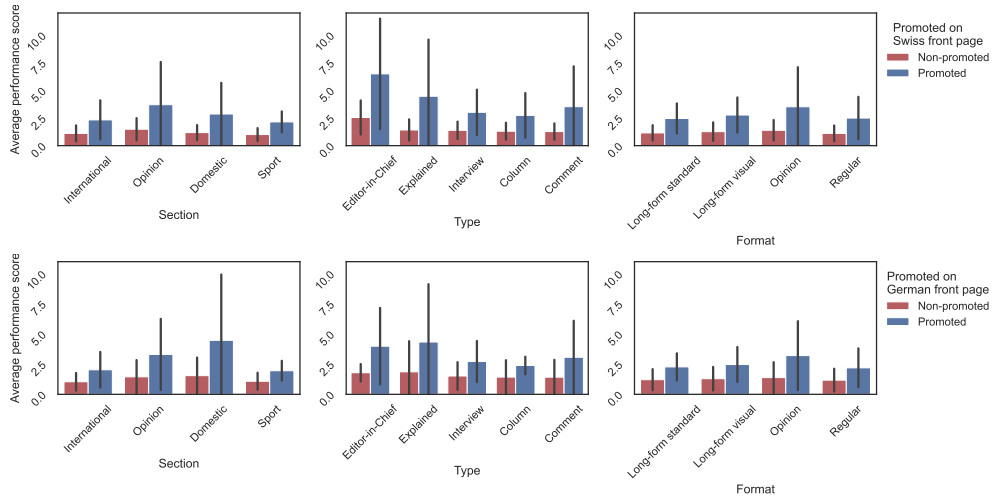


Figure 5 Average performance score by the section, type, and format across promoted and non-promoted

items. Error bars are standard deviations from 1000 bootstrap runs. Top: Swiss front page. Bottom: German front page.

- **Random policy:** This policy selects items to promote at random. By the properties of the hypergeometric distribution (i.e., sampling a fixed number of “successes” without replacement), the probability that a randomly selected item is promoted is $C_{tm}/|\mathcal{I}_t|$, where C_{tm} is the number of items promoted on channel m in time period t and \mathcal{I}_t is the pool of relevant

items to choose from. A random policy is then simply given by

$$d_m(i \in \mathcal{I}_t; \mathcal{C}_{tm}) = \begin{cases} 1, & \text{with probability } C_{tm}/|\mathcal{I}_t|, \\ 0, & \text{with probability } 1 - C_{tm}/|\mathcal{I}_t|. \end{cases} \quad (23)$$

This represents a worst-case non-contextual policy. We include it to measure the extent to which the current practice is optimizing and to quantify the economic opportunity cost had a randomized experiment been run.

- **Autoregressive policy:** Discussions with the company revealed that current practice is primarily based on item performance in the immediate past. A data-driven heuristic of this practice is to, per time period, promote the items that had the largest performance scores in the previous time period. Let $Y_{i,t-1}$ be the reward of item i in time period $t-1$ and let $Y_{t-1}^{(C_{tm})}$ be C_{tm} -th largest reward among the items at that time. An autoregressive policy can then be written as

$$d_m(i \in \mathcal{I}_t; Y_{i,t-1}) = \begin{cases} 1, & \text{if } Y_{i,t-1} \geq Y_{t-1}^{(C_{tm})}, \\ 0, & \text{else,} \end{cases} \quad (24)$$

where always promoting the same items is avoided by that new time periods contain new items. Including this policy allows us to measure whether replacing current manual practice with a data-driven algorithm can *by itself* improve promotion decisions, where we isolate the performance contribution of being data-driven from other factors (e.g., using machine learning for prediction, using an allocation mechanism theoretically known to be better) by intentionally designing this policy's allocation mechanism as a heuristic of current practice, where we consider the utilized context to be one of the main signals that is used in current practice.

- **Non-incremental policy:** This policy promotes the items with the largest predicted rewards if they would be promoted. The policy can be written as

$$d_m(\mathbf{X}_{it}) = \begin{cases} 1 & \text{if } \hat{Y}(\mathbf{X}_{it}, A_{itm} = 1, \mathbf{A}_{it}^{-m}) \geq \hat{Y}_t^{(C_{tm})}, \\ 0 & \text{else.} \end{cases} \quad (25)$$

where $\hat{Y}^{(C_{tm})}$ C_{tm} -th largest predicted reward under promotion among the items in time period t . The policy is identical to the optimal policy in all aspects (e.g., it uses a machine learning model for prediction based on the full context), except that its allocation mechanism

is based on outcome prediction, not incremental effects. This type of policy is commonly used in practice but is generally sub-optimal (e.g., Ascarza 2018, Hitsch et al. 2024), as it neglects that the decisions with the greater outcomes from promotion are not necessarily those whose outcome increases the most from promotion. Including this into a policy lets us isolate the part of the performance improvement that is caused by making decisions based on predicted incremental effects instead of predicted potential outcomes.

4.4.2. Performance and revenue metrics. We compare the performance of the policies in terms of their performance gain over current practice (i.e., the behavior policy) at our partner company. We define the performance gain of a policy d_m as the relative improvement in expected reward (value) if that policy would have been implemented instead of the behavior policy, that is,

$$G_V(d_m) = \frac{V(d_m) - V(\pi_m)}{V(\pi_m)}. \quad (26)$$

Because the true population value of a policy is always unknown, we calculate the gain on data in terms of the sample estimates on the hold-out test set. We obtain the value of the historical decision under current practice, $\widehat{V}(\pi)$, as simply the sample average outcome in the test set, and for each evaluated policy d_m , we estimate its counterfactual value, $\widehat{V}(d_m)$, had it been implemented instead on the test set, by applying the value estimator in Eq. (13). Note that the value estimates of the different policies are derived from the same fitted and hyperparameter-tuned nuisance models. Hence, any difference in performance can only be attributed to differences in treatment policies.

We also consider the revenue implications of the promotion policies. Let $R(\cdot)$ be the total revenue from adopting a promotions policy. Then

$$R(d_m) - R(\pi_m) = R(\pi_m) \frac{dG_R}{dG_V} G_V(d_m) \quad (27)$$

measures the dollar gain of a new promotions policy d_m over current practice π_m , where

$$G_R(d_m) = \frac{R(d_m) - R(\pi_m)}{R(\pi_m)}. \quad (28)$$

is its relative revenue gain, and dG_R/dG_V is the rate of return in revenue to relative improvements in promotions policy performance over current practice. To facilitate a fair comparison, we assume this rate is independent across policies, and therefore drop the policy dependence in the notation.

The performance of promotion policies also depends on the capacity constraint. To further ensure a fair comparison between the policies and current practice, we set the capacity constraint C_{tm} to the same as in the historical data, for both channels and every time period. This way, any

performance or revenue difference cannot be explained by that policies promote more or fewer but only which items they promote. Later, we relax this constraint to study the returns to increasing the number of promotions. The idea behind this is that the adoption of algorithmic decision-policies in business is often motivated by a potential for efficiency gains.

4.4.3. Comparison of promotion policies. Table 1 reports the results for performance and revenue gains. For the latter, we base our calculations on public financial figures indicating that 89% of the USD 285 million total revenue of our partner company in 2022 came from ads, sales, and subscriptions (DigiDay 2018, Center 2021, Statista 2022, DigiDay 2023) and internal analyses suggesting that each 1% gain in the performance score of traffic, engagement and subscriptions leads to a 0.1–0.5% increase in revenue. To estimate the revenue gain of a new policy according to Eq. (27), we thus set $R(\pi_m) = 285 \cdot 0.89 = 254$ million USD, dG_R/dG_V between 0.001 and 0.005, and $G_V(d_m)$ according to Eq. (26). We used the first 3 weeks of historical data as the training set for the estimation and the last 3 weeks as the test set for off-policy evaluation, for which we report all results.

We have three main findings. First, our optimal policy performs best. It leads to a 5.21% (Swiss front page) and 2.25% (Germany front page) expected gain in the performance score of traffic, engagement, and subscriptions, a statistically significant improvement over current practice at the 95% confidence level. This implies an economically significant increase of USD 1.32–6.62 million in revenue. Effect sizes of digital marketing interventions are typically very small and hard to detect, so the statistical and economic significance is practically important.

Second, our evaluation of the alternative policies shows the comparative value of our framework. Here, the non-incremental policy is estimated to result in 0.8–5.41 million USD less promotional revenue across the two channels than the optimal policy, thus quantifying the sub-optimality of outcome prediction approach from a business perspective.

The autoregressive policy, which heuristically mimics the editors’ decision process, outperforms them. This suggests that data-driven policies not based on machine learning or causal inference, but simple descriptive statistics (in this case, the observed outcomes from promotions in the past), can outperform expert decision-makers.

As hypothesized, the random policy does not beat the current practice. This confirms that the editors were (imperfectly) optimizing, and empirically shows the need to account for selection bias in the off-policy learning, as per our framework. The revenue implications of the random policy further suggest that running an A/B test would have cost the company up to USD 14 million

(9.75 for the Swiss front page plus 4.24 for the German front page) due to a loss in promotional performance. This corresponds to 6% of total annual revenue from ads or subscriptions according to our figures, and is several orders of magnitude greater than the revenue gain of the optimal policy. Overall, this adds strength to our framework based on historical data.

Third, all improvements are larger for the Swiss front page. Put differently, the company’s current practice is estimated to be closer to optimum for the German market than for the Swiss market. Discussions with our partner company revealed that this may be because the newspaper only recently expanded to Germany. Therefore, the newspaper has a relatively smaller market share and a more homogenous audience in Germany than in Switzerland, where the newspaper is long-established and has a broader audience. As such, it is more difficult for the editors to satisfy their Swiss users’ preferences with audience-wide promotions. This finding points towards an inherent challenge in improving audience-wide promotions with increasing reach.

Table 1 Performance and revenue gains on the test set (out-of-sample).

Policy	<i>Swiss front page</i>		<i>German front page</i>	
	Performance gain	Revenue gain (million USD)	Performance gain	Revenue gain (million USD)
Random	−7.69%	−9.75 to −1.95	−3.34%	−4.24 to −0.85
Autoregressive	1.94%	0.49 to 2.48	0.85%	0.23 to 1.07
Non-incremental	3.26%	0.83 to 4.13	1.01%	0.26 to 1.28
Optimal	5.21%	1.32 to 6.61	2.25%	0.57 to 2.85

Notes. Performance gain = Percentage increase in the performance score of traffic, engagement, and conversions on average across content and time periods (cf. Eq. (26)). Revenue gain = Absolute increase in million USD implied by the performance gain (cf. Eq. (27)), given a base annual profit of 253 million USD revenue from ads, sales, and subscription and a conservative revenue elasticity of promotions policy performance of 0.1–0.5%. Larger is better. Best policy in bold.

4.4.4. Returns to changing the capacity constraint. We now examine the returns in promotional value to adapting the capacity constraint of the number of items that can be promoted on a channel at a time. We evaluate two scenarios: (1) where we set the capacity to the average number of promotions per channel and time period in the historical data; and (2) where we set the capacity to the maximum number of items that is possible to show on a front page at a time. Scenario (1) isolates the returns from using a “fixed” constraint, as the sum of the time-averaged number of promotions over time is, by construction, equal to the sum of the dynamically changing number of promotions over time. Scenario (2) quantifies the returns to making the most use of the algorithmic decision policies, as a common motivation for their use in businesses is to scale allocation decisions beyond what is feasible manually.

The results are reported in Table 2. We omit the random policy as, by design, it has no returns to increasing capacity. For both markets, we find that fixing the capacity constraint per scenario (1)

leads to a lower performance and revenue gain than using the original dynamic capacity constraint. If the total number of promotions over time cannot be increased, then this speaks in favor of adapting the number of promotions per time period and channel to the dynamics in the data or, equivalently, to what was historically chosen by the experts. As for scenario (2), we find that, if it is instead possible to promote more than in the past, then promoting the maximum amount of items does lead to additional incremental performance and revenue. For the Swiss market, in particular, it increases the performance gain from 5.21% to 13.21%, and the revenue gain from 1.32–6.61 to 3.35–16.75 million USD. This confirms the returns to making full use of the capabilities of algorithmic allocation policies.

Table 2 Performance and revenue returns to increasing promotions capacity (out-of-sample).

Policy	<i>Swiss front page</i>		<i>German front page</i>	
	Performance gain	Revenue gain (million USD)	Performance gain	Revenue gain (million USD)
<i>$C_{tm} = 30$ promotions</i>				
Autoregressive	1.32%	0.33 to 1.67	0.62%	0.16 to 0.79
Non-incremental	2.56%	0.65 to 3.25	0.70%	0.18 to 0.89
Optimal	4.35%	1.19 to 5.52	1.94%	0.49 to 2.46
<i>$C_{tm} = 60$ promotions</i>				
Autoregressive	8.78%	2.23 to 11.14	2.87%	0.73 to 3.64
Non-incremental	10.33%	2.62 to 13.10	3.19%	0.81 to 4.05
Optimal	13.21%	3.35 to 16.75	5.05%	1.28 to 6.40

Notes. Performance gain = Percentage increase in the performance score of traffic, engagement, and conversions on average across content and time periods (cf. Eq. (26)). Revenue gain = Absolute increase in million USD implied by the performance gain (cf. Eq. (27)), given a base annual profit of 253 million USD revenue from ads, sales, and subscription and a conservative revenue elasticity of promotions policy performance of 0.1–0.5%. Larger is better. Best policy in bold.

4.5. Explaining Policy Improvement

We now seek to understand the improvement of the optimal policy over current practice. To do so, we analyze different sources of heterogeneity in policy outcomes that contribute to the overall gain. We focus on heterogeneity across time, across content items, and across the decisions.

4.5.1. Temporal heterogeneity. Our results across hour of day and day of the week are shown in Fig. 6. We find that the optimal policy performs consistently best. Moreover, we observe that the average value under our optimal policy is characterized by similar dynamics as the average value of the behavior policy, which adds to the credibility of our counterfactual results. For example, the overall news consumption is particularly pronounced in the early morning, and we thus also expect a large gain from optimizing content promotions for the early morning, which is confirmed by our results.

4.5.2. Cross-sectional heterogeneity across items. We also find that, under the optimal policy, the average performance score varies by the section, type, and format of content (Fig. 7). The value is slightly higher for opinion and explanatory articles and substantially larger for articles written by the editor-in-chief. The results are similar across the two front pages but different in magnitude. Overall, promoting items has a positive effect on the average performance of the corresponding items, but the magnitude depends both on when the items are promoted and their characteristics.

4.5.3. Promotional heterogeneity. The optimal policy outperforms current practice by promoting different items. To further understand where these improvements stem from, we examine the heterogeneity in the categories of content that tend to be promoted by the optimal policy vs. the behavior policy.

Fig. 8 compares the share of items that are promoted by the optimal policy vs. the behavior policy per section, type, and format. Similar to current practice, the optimal policy promotes items from different categories at different rates. For example, our results suggest that our partner company should promote relatively more items belonging to the “Opinion” section but fewer belonging to the “International” section. We also find heterogeneity concerning what types of items should be promoted on the two front pages. Historically, about 40% of the articles written by the editor-in-

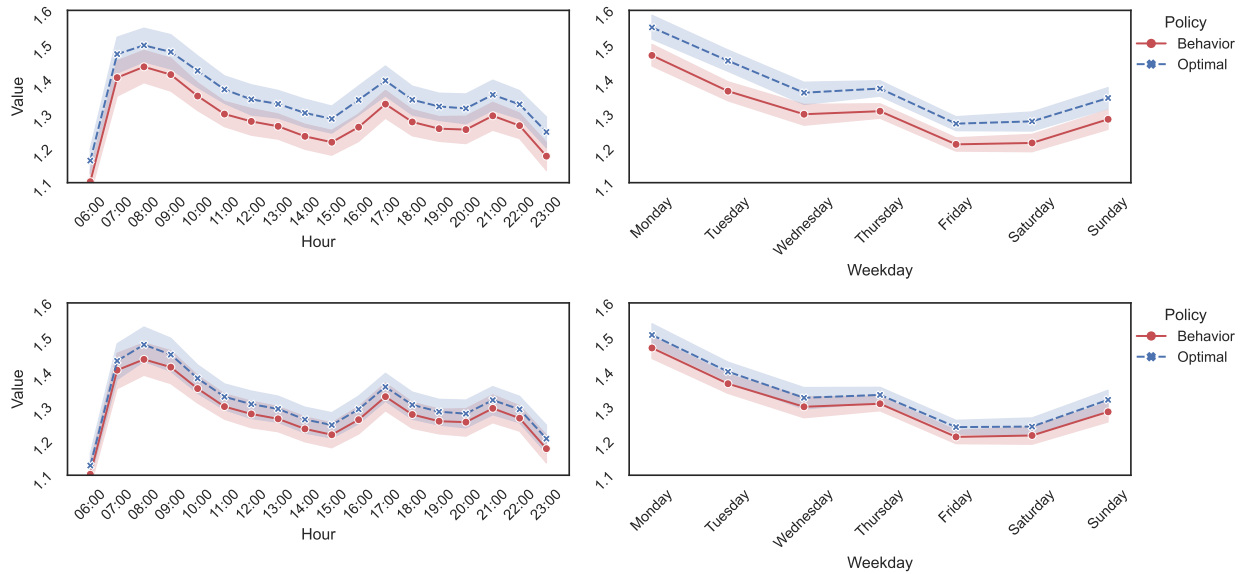


Figure 6 Value of the behavior policy (current practice) and our optimal policy by hour and day of the week.

Shaded regions are 95% confidence intervals across 1000 bootstrap runs. Top: Swiss front page. Bottom: German front page.

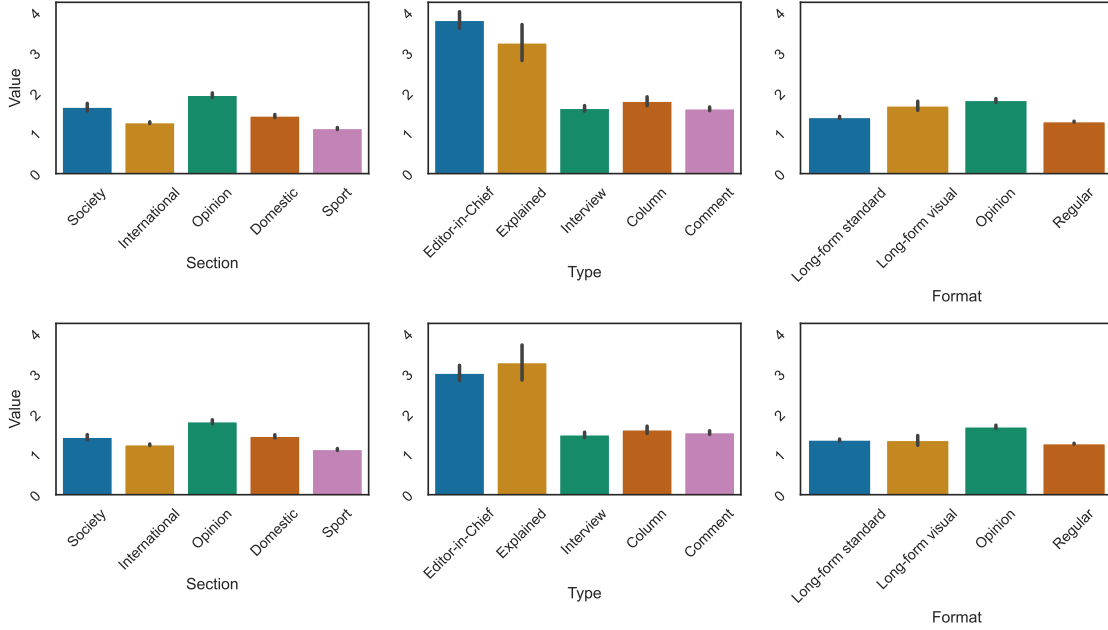


Figure 7 Value of optimal policy by the sections, types, and formats of items. Error bars are 95% confidence intervals across 1000 bootstrap runs. Top: Swiss front page. Bottom: German front page.

chief were promoted. For the Swiss but not the German front page, the optimal policy recommends the rate should increase to around 95%. A possible explanation stems from that the newspaper is based in Switzerland. Thus, articles by the editor-in-chief may resonate more with the Swiss market. This is also supported by Fig. 8, which shows that those articles have a substantially higher CATE for the Swiss front page.

Previous research has found that the sentiment of content is a strong driver of views and engagement (Berger and Milkman 2012, Upworthy 2012, Robertson et al. 2023). Here, we find the optimal policy promotes positive, neutral, and negative content at a similar rate (top row of Fig. 9). The result for the optimal policy is explained by that the estimated CATE of promotion is roughly the same for different content (bottom row of Fig. 9). We later discuss this finding in Sec. 5.

4.6. Drivers of Effect Heterogeneity and Optimal Decisions

4.6.1. Promotion effect heterogeneity. We now aim to find which context covariates are moderators of the heterogeneity promotion effects and, therefore, explain how the optimal policy arrives at its decisions. We consider two approaches: (i) *SHAP values*: We use this approach to check the variable importance with respect to non-linear CATE heterogeneity. (ii) *Post-Lasso inference*: For valid statistical inference with respect to marginal CATE heterogeneity.

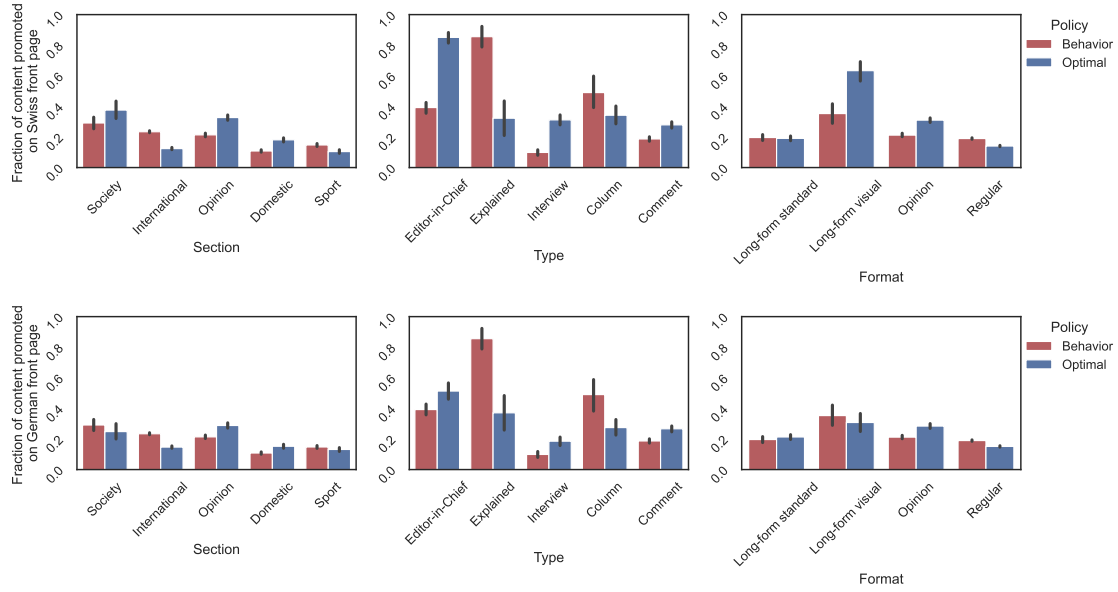


Figure 8 Share of items promoted by optimal policy vs. behavior policy by content category. Error bars are 95% confidence intervals across 1000 bootstrap runs. Top: Swiss front page. Bottom: German front page.

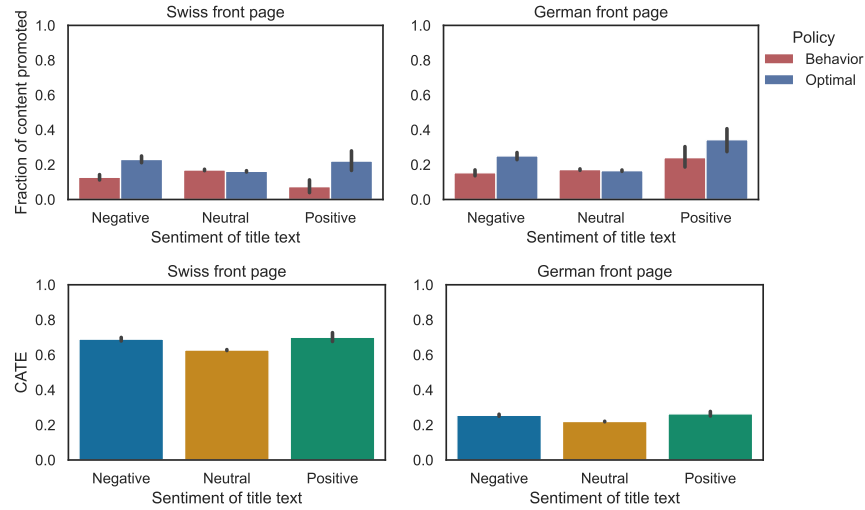


Figure 9 Top: Share of content promoted by optimal policy vs. behavior policy by sentiment of content title. Bottom: Estimated CATE by sentiment of content title. Error bars are 95% confidence intervals across 1000 bootstrap runs.

Our first approach is to calculate the SHAP values of the CATE function. The SHAP value of a covariate measures its contribution to a prediction relative to the average prediction of the model (Lundberg and Lee 2017). A benefit of SHAP values in our work is that they directly explain the logic of the estimated CATE model by accounting for non-linearities in the treatment

effect heterogeneity; however, it does not allow for valid statistical inference on the discovered heterogeneity. Hence, we use the SHAP values to provide initial findings of the heterogeneity that can be explained rigorously by the post-Lasso inference method.

The results are shown in Fig. 10. The covariates are ordered from top to bottom in terms of their mean absolute SHAP values, and the dots represent the SHAP values of the covariates in terms of explaining heterogeneity in the CATE. The horizontal spread of the dots for a given covariate corresponds to the variance in the CATE that is explained by that covariate, relative to the average CATE (i.e., the average treatment effect). Comparing the SHAP values for the Swiss and the German front pages, we find that 14 out of the top-15 covariates are the same, but that their relative importance differs slightly. In particular, average engagement time, average scroll depth, and previous email newsletter promotions are among the top-5 most important covariates in terms of explaining the heterogeneity in the CATE of promotion on both front pages, but the relative importance of the rest differs. Overall, this demonstrates the importance of accounting for between-channel heterogeneity in the moderators of promotional effectiveness and, therefore, that promotion decisions should be based on different contextual information for the two channels.

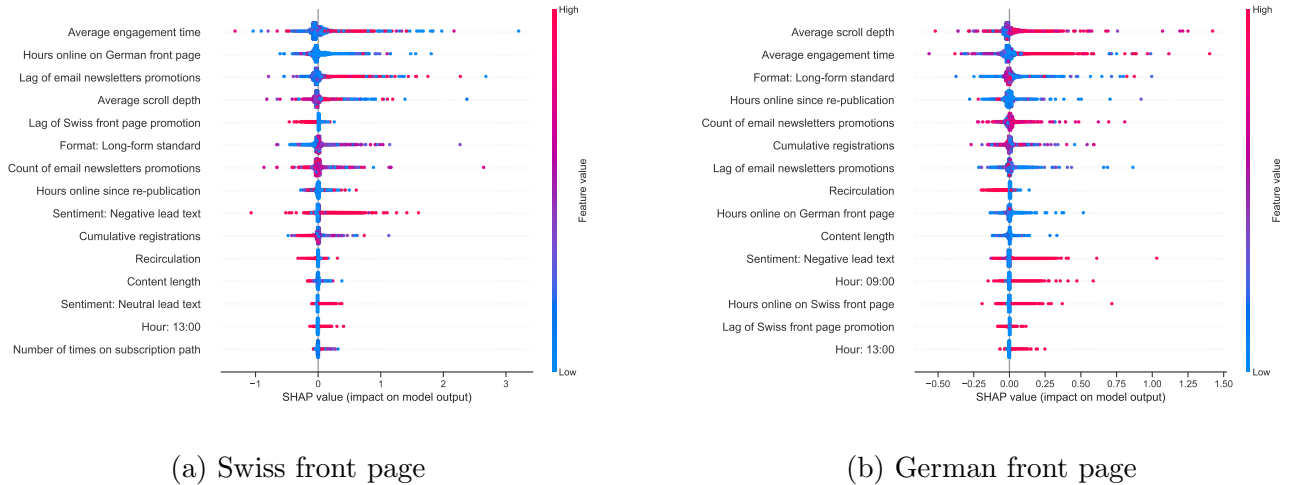


Figure 10 SHAP values for the top-15 covariates with the largest feature importance on the predictions of the CATE function (in descending order).

For our post-Lasso inference approach, we adapt the method in (Chernozhukov et al. 2018b, Semenova et al. 2023), which estimates the best (lower-dimensional) linear projection (BLP) of the CATE. This allows for discovering marginal heterogeneity in the CATE with respect to covariates, even when the true CATE is non-linear and unknown (Chernozhukov et al. 2018b). A straightforward approach to estimate the BLP of the CATE is to regress the final CATE estimates on all

context covariates subject to a Lasso penalty. However, Lasso induces regularization bias in the point estimates of the regression coefficient on the selected covariates and does not allow for valid post-selection inference. To address this, we combine the Lasso-regularized estimation of the BLP of the CATE with post-Lasso inference in a sample-splitting procedure (for theoretical details; see, e.g., Wasserman and Roeder 2009, Belloni and Chernozhukov 2013, Belloni et al. 2014, Zhao et al. 2021).

Our procedure consists of the following steps.

1. We randomly partition the training data \mathcal{T} into two mutually exclusive splits \mathcal{T}_1 and \mathcal{T}_2 .
2. On split \mathcal{T}_1 , we fit a Lasso regularized linear model where the CATE function estimates are the dependent variable and the context covariates are the predictors. For estimation, we use a suitable optimization algorithm (i.e., coordinate descent) to minimize the mean square error between the CATE estimates and the linear model subject to the Lasso regularization penalty,

$$(\hat{\alpha}_m, \hat{\beta}_m^0) = \arg \min_{\alpha_m, \tau_m} \frac{1}{|\mathcal{T}_1|} \sum_{t \in \mathcal{T}_1} \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} \left(\hat{\tau}_m(\mathbf{x}_{it}) - \alpha_m - \boldsymbol{\beta}_m^\top \mathbf{X}_{it} \right)^2 + \lambda \sum_{k=1}^p |\beta_k|. \quad (29)$$

Here, $\lambda \in [0, \infty]$ is a tuning parameter for the amount of regularization and $\hat{\beta}_m^0 \in \mathbb{R}^l$, $l \leq \dim(\mathbf{X}) = p$ are the non-zero parameter estimates for the selected covariates $\mathbf{X}^0 := \{X_k : \hat{\beta}_k \neq 0\} = \{X_k \in \mathbf{X} : \hat{\tau}_m(\mathbf{X}) - \hat{\tau}_m(\{\mathbf{X} \setminus X_k\}) \neq 0\}$.

3. On the other split \mathcal{T}_2 , regress the CATE estimates on the Lasso-selected covariates,

$$\hat{\tau}_m(\mathbf{X}_{itm}^0) = \alpha_m + \boldsymbol{\xi}_m^\top \mathbf{X}_{itm}^0 + \epsilon_{itm} \quad (30)$$

using ordinary least squares (OLS) for parameter estimation, $\hat{\boldsymbol{\xi}}_m = (\mathbf{X}_m^{0\top} \mathbf{X}_m^0)^{-1} \mathbf{X}_m^{0\top} \hat{\boldsymbol{\tau}}_m$, and HC2 robust standard errors for inference.

It follows that the least-squares parameter estimates measure marginal effects of the Lasso-selected covariates with respect to treatment effect heterogeneity.⁸ See Appendix D.3 for details on the cross-validation and hyperparameter tuning for the Lasso regularization.

The results from the post-Lasso inference approach are shown in Table 3. First, within channels, the marginal effects of covariates on heterogeneity in the CATE vary substantially. On average, and all else equal, content with a one standard deviation higher average engagement time (relative to the number of characters of the content) have a get a 0.024 and 0.014 points greater return

⁸ If the Lasso-regularization selects the true covariate set, then the estimates equal those of the oracle estimator, and, if the Lasso-regularization does not select the true covariates, then the estimates are still less biased and faster converging than standard Lasso estimates. This holds in finite-samples with high-dimensional controls, non-Gaussian and heteroskedastic errors, and even if the underlying true CATE is non-parametric (Belloni and Chernozhukov 2013, Belloni et al. 2014), all of which may apply to our framework.

to promotion on the Swiss and German front page, respectively. In contrast, content that has already been promoted in email newsletter tend to benefit less from promotion than those items who have not already been distributed in that channel, which may be explained by many users have already seen and engaged with the promoted item. Among past performance indicators (up until the start of the current time period), we find that more views, longer engagement time, and more registrations predict a higher incremental effect from promotion. Among content types, articles written by the editor-in-chief benefit the most from promotion on the Swiss front page, but are not predicted to benefit from promotion on the German front page. This is consistent with our previous results that our optimal policy promotes articles by the editor-in-chief relatively more often to the Swiss front page than the German front page. Second, we also find some differences in marginal treatment effect heterogeneity across the channels. For example, articles in a visual format are predicted to have a lower CATE for the Swiss front page and a higher CATE for the German front page, relative to content in standard format. Additional results evaluating the performance in the first-stage Lasso are in Appendix 4.6.2.

Overall, we find that covariates with large feature importance in our analysis based on the SHAP value method (e.g., average engagement time and average scroll depth) tend to have a comparatively large absolute coefficient in the post-Lasso inference approach. Therefore, the results from both approaches for post-hoc explainability appear to be largely consistent.

4.6.2. Variable selection for optimal promotion decisions. We extend our post-Lasso inference procedure for the CATE to variable selection for the optimal policy. Our idea is to leverage that (i) the post-Lasso inference procedure allows for valid inference on which covariates contribute to heterogeneity in the CATE, and (ii) the optimal policy is uniquely determined by heterogeneity in the CATE; therefore, (iii) the Lasso-selected covariates that have true non-zero marginal effects on the variation in the CATE are sufficient for the optimal policy. As such, we discover which covariates these are by conducting the following two-sided hypothesis test for covariate X_j in the channel-specific Lasso-selected set \mathbf{X}_m^0 :

$$H_0^j : \xi_{jm} = 0 \quad \text{vs.} \quad H_A^j : \xi_{jm} \neq 0. \quad (31)$$

The set of covariates with a statistically significant contribution to CATE heterogeneity is given by

$$\{\mathbf{X}_m^\alpha\} := \{X_j \subseteq \mathbf{X}_m^0 : H_0^j \text{ is rejected at some significance level } \alpha/l\}, \quad (32)$$

Covariate	CATE for Swiss front page			CATE for German front page		
	Coef.	SE	p-value	Coef.	SE	p-value
<i>Past performance indicators:</i>						
Article views	0.033	0.013	0.009	0.011	0.004	0.005
Average engagement time	0.024	0.007	0.001	0.014	0.004	0.000
Average scroll depth	−0.033	0.002	0.000	0.011	0.001	0.000
Recirculation	−0.009	0.001	0.000	−0.005	0.001	0.000
Cumulative registrations	0.059	0.007	0.000	0.023	0.002	0.000
Number of times on subscription path	0.004	0.003	0.154	0.001	0.001	0.292
<i>Time-invariant content characteristics:</i>						
Content length	−0.002	0.001	0.262	−0.004	0.001	0.000
Section: Business & finance	−0.062	0.007	0.000	−0.032	0.006	0.000
Section: Celebrities & events	0.066	0.012	0.000	−0.005	0.006	0.433
Section: Culture	−0.077	0.008	0.000	−0.035	0.006	0.000
Section: Domestic	−0.028	0.009	0.003	−0.029	0.006	0.000
Section: Education	0.097	0.016	0.000	0.036	0.009	0.000
Section: International	−0.066	0.007	0.000	−0.024	0.006	0.000
Section: International politics	−0.047	0.010	0.000	−0.009	0.009	0.292
Section: NZZ in English	0.053	0.012	0.000	0.059	0.008	0.000
Section: Opinion	−0.170	0.011	0.000	−0.022	0.008	0.005
Section: Science & technology	−0.051	0.010	0.000	−0.024	0.007	0.001
Section: Society	−0.002	0.035	0.950	−0.032	0.010	0.002
Section: Sport	−0.034	0.010	0.000	−0.026	0.006	0.000
Section: Visuals	−0.077	0.011	0.000	−0.077	0.010	0.000
Section: Zurich	−0.055	0.008	0.000	−0.030	0.006	0.000
Type: Breaking news	−0.175	0.006	0.000	−0.022	0.003	0.000
Type: Column	0.025	0.013	0.061	0.016	0.006	0.011
Type: Comment	0.088	0.010	0.000	0.034	0.005	0.000
Type: Editor-in-Chief	0.431	0.024	0.000	—	—	—
Type: Explained	0.106	0.022	0.000	0.008	0.007	0.295
Type: Guest comment	0.115	0.011	0.000	—	—	—
Type: Interview	0.052	0.011	0.000	0.014	0.004	0.000
Type: News in brief	−0.040	0.017	0.021	−0.008	0.004	0.054
Format: Long-form standard	−0.038	0.005	0.000	0.002	0.003	0.604
Format: Long-form visual	−0.040	0.011	0.000	0.013	0.006	0.022
Format: Opinion	0.023	0.009	0.017	0.018	0.005	0.000
Sentiment: Negative lead text	0.113	0.006	0.000	0.016	0.003	0.000
Sentiment: Positive lead text	—	—	—	—	—	—
Sentiment: Negative SEO title	0.029	0.008	0.001	0.007	0.003	0.010
Sentiment: Positive SEO title	0.035	0.018	0.053	0.006	0.004	0.075
Sentiment: Negative title	−0.023	0.006	0.000	0.004	0.002	0.073
Sentiment: Positive title	−0.031	0.009	0.001	—	—	—
<i>Time information:</i>						
Hours online on Swiss front page	0.002	0.001	0.000	0.001	0.000	0.000
Hours online on German front page	−0.001	0.001	0.189	0.001	0.000	0.000
Hours online since publication	−0.005	0.002	0.040	−0.008	0.001	0.000
<i>Past promotions:</i>						
Count of social media promotions	0.040	0.007	0.000	−0.026	0.001	0.000
Count of email newsletters promotions	−0.010	0.001	0.000	−0.007	0.001	0.000
Count of push notification promotions	−0.047	0.007	0.000	0.020	0.001	0.000
Lag of social media promotions	0.005	0.001	0.000	0.004	0.001	0.001
Lag of email newsletters promotions	0.006	0.001	0.000	0.002	0.001	0.031
Lag of push notification promotion	0.003	0.002	0.132	0.002	0.001	0.095
Lag of Swiss front page promotion	−0.017	0.002	0.000	−0.011	0.001	0.000
Lag of German front page promotion	−0.069	0.002	0.000	0.000	0.001	0.993
Hour fixed effects	Yes			Yes		
Weekday fixed effects	Yes			Yes		
R-squared	0.195			0.151		
Adjusted R-squared	0.194			0.150		
AIC	14 760			−51 260		
BIC	15 360			−50 690		

Coef.: point estimate of marginal effect. SE: robust (HC2) standard errors.

Table 3 Marginal effects of covariates (using pre-treatment values) for explaining heterogeneity in the CATE.

The estimates are obtained from the BLP of the CATE function estimated with OLS in our post-Lasso inference procedure. Cells with a “—” denote that the corresponding covariate was dropped by the first-stage Lasso regularization. Continuous covariates are z -standardized to zero mean and unit variance for better interpretability.

where α/l is the significance level after Bonferroni correcting for $l = \dim(\mathbf{X}_m^0)$ tests. Our way of combining post-Lasso inference with a multiple testing correction is similar to the method of Pelger and Zou (2022), but whereas they consider the problem of explaining a high-dimensional linear panel data model, we consider selecting a sufficient set of covariates for an optimal decision policy. Nonetheless, the method of Pelger and Zou (2022) may be viewed as providing the theoretical basis of our approach.

Table 3 shows that the Lasso-regularization selected $l = 48$ and $l = 45$ CATE-relevant covariates for the Swiss and German front pages, of which only those with a post-selection p-value of at most $\alpha/l \approx 0.001$ should be included in the further reduced covariate sets in Eq. (32) after applying the Bonferroni correction. This leaves us with 36 out of the 48 covariates that were CATE-relevant for Swiss front page, and 27 out of the 45 covariates that we CATE-relevant for German front page.

Using the variables selected for either front page, we re-estimate the orthogonal random forests for the CATE model h_m separately for the channels, yet again using our sample-split cross-fitting described in Sec. 3.6.4. Just as for all other policies, we off-policy evaluate the value of the resulting *reduced optimal policy* by applying the AIPW estimator in Eq. (13) to the test set, where, also as before, we use the same fitted nuisance models for rewards and propensity scores depending on the full context covariate set, the optimal policy decisions now stem from CATE estimates derived from the new CATE models that only used the reduced covariate sets, i.e., $h_m(\mathbf{X}_m^\alpha)$. Thus, our reduced optimal policy only differs from our “standard” optimal policy in that it only leverages the selected subset of context covariates that are CATE-relevant for the channel.

Table 4 reports the promotional performance and revenue gains of the reduced optimal policy. First, and in line with our expectations, the reduced optimal policy leads to substantial incremental promotions performance and revenue over the the behavior policy (i.e., current practice). Second, its expected performance and revenue gains is also greater than that of our optimal policy that is based on all context covariates. However, after accounting for uncertainty in terms of the confidence intervals around these estimates we see that the value of the policies is on average statistically indistinguishable at all hours of the day and for each day per week (Fig. 11). This shows that there is no significant loss in the performance or revenue of an decision policy based on the CATE when it is reduced to only those contexts that informs its treatment assignment mechanism. In a broader sense, this also demonstrates that even a subset of the data collected by our partner company is sufficient to make optimal promotion decisions. Taking our partner company as an example, the managerial implication is that the dashboard would only need to show data on at

most three-quarters of all context covariates, and that is possible to efficiently learn which these are using our data-driven procedures. This could support the editors’ decision-making by reducing the amount of information they need to process for making promotion decisions.

Table 4 Policy value estimates on the test set.

Policy	<i>Swiss front page</i>		<i>German front page</i>	
	Performance gain	Revenue gain (million USD)	Performance gain	Revenue gain (million USD)
Optimal policy	5.21%	1.32 to 6.61	2.25%	0.57 to 2.85
Reduced optimal policy	6.76%	1.71 to 8.57	2.95%	0.75 to 3.74

Notes. Performance gain = Percentage increase in the performance score of traffic, engagement, and conversions on average across content and time periods (cf. Eq. (26)). Revenue gain = Absolute increase in million USD implied by the performance gain (cf. Eq. (27)), given a base annual profit of 253 million USD revenue from ads, sales, and subscription and a conservative revenue elasticity of promotions policy performance of 0.1–0.5%. Optimal policy: Our optimal policy when the predicted CATEs are based on all context covariates. Reduced optimal policy: The optimal policy when the predicted CATEs are based only on the context covariates selected by our post-Lasso hypothesis testing inference. Larger is better. Best policy in bold.

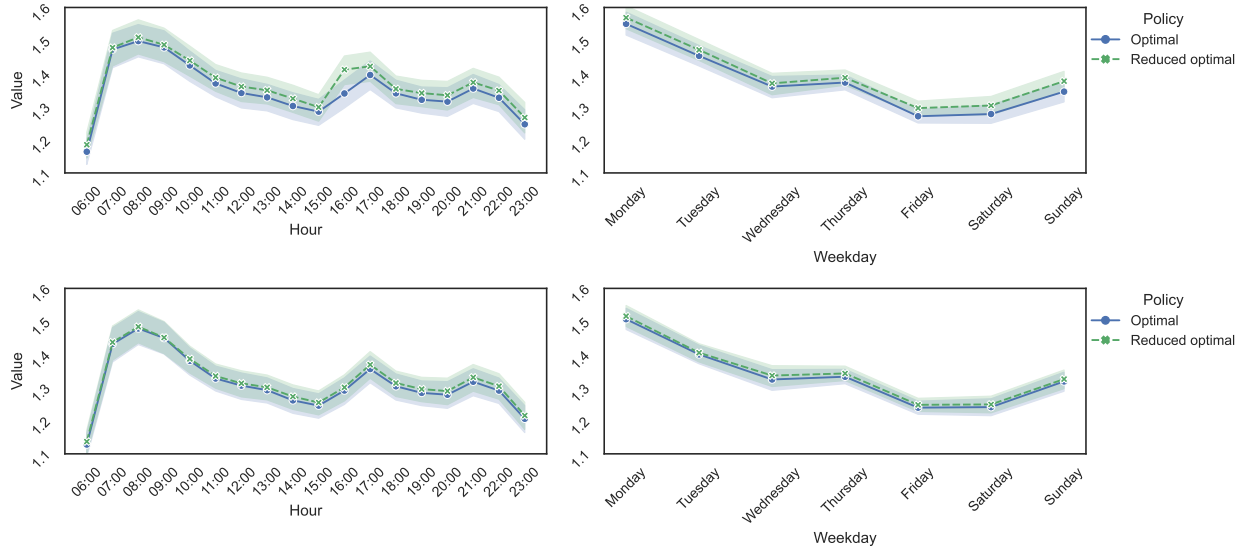


Figure 11 Value of the optimal policy and the reduced optimal policy over time. Shaded regions are 95% confidence intervals across 1000 bootstrap runs. Top: Swiss front page. Bottom: German front page.

5. Discussion

In this paper, we presented an off-policy learning framework for optimizing audience-wide content promotions across digital distribution channels, a common decision problem faced by content publishers and producers (e.g., newspapers, content creators), who must choose which of their content to promote to their entire audience such that the expected reward (e.g., traffic, engagement, or conversions) is maximized.

Our work contributes in several novel ways. Previous research in marketing and machine learning has studied on-policy learning and outcome prediction (e.g., adaptive experiments and bandit algorithms) for content personalization (e.g., Agarwal et al. 2009, Hauser et al. 2009, Kale et al. 2010, Li et al. 2010, Garcin et al. 2013, Urban et al. 2014, Schwartz et al. 2017). Another line of work has shown that personalized marketing decisions, whether online or offline, should be based on incremental effects, not outcome prediction (Bodapati 2008, Ascarza 2018, Lemmens and Gupta 2020, Hitsch et al. 2024). An emerging stream of research has adopted this incremental effects approach with off-policy learning to optimize personalization in various marketing contexts (e.g., Simester et al. 2020a, Liu 2022, Ellickson et al. 2023, Yoganarasimhan et al. 2023, Hitsch et al. 2024, Huang and Ascarza 2024, Yang et al. 2024). We contribute by studying the common but relatively understudied and non-personalized decision problem of selecting which content to promote to an entire audience and, for this, provide a framework leveraging off-policy learning and the incremental effects approach.

Unlike previous research in the aforementioned streams, our framework is designed for use with historical data. This is particularly advantageous for optimizing audience-wide promotion strategies, where randomizing content to promotion as in A/B test is typically impractical, costly, and risky. Our approach of using a model of the historical data-generating process to enable identification, may be of interest in other marketing decision-problems where randomization is infeasible but rich historical data is available. Previous research (Lada et al. 2019, e.g.,) has demonstrated the value of historical data for learning decision-policies based on causal effects, but only for the personalized setting and without connecting it to identification.

We enhance the practical utility of our framework by integrating our model with a causal machine learning procedure for estimation, which we show is unbiased for the optimal promotion policy under weak assumptions, selection bias, or reward misspecification, thereby ensuring its robustness for use with historical data. We combine the estimation procedure with a post-Lasso inference approach to discover drivers of the promotional effectiveness, and combine it with multiple testing to select which variables in the high-dimensional contexts are sufficient to retain the optimal policy’s performance. Overall, our framework addresses real-world constraints of randomized experiments and offers a robust and cost-effective solution to optimize content promotion strategies in a data-driven manner.

We also introduce an off-policy evaluation approach similar to ablation studies, which allows for causal comparisons of different marketing decision strategies. Unlike previous research that

evaluates the outcomes of a fixed optimal policy with different machine learning models (e.g., Hitsch et al. 2024, Smith et al. 2022), our approach quantifies how altering the treatment assignment mechanism impacts policy outcomes, holding the machine learning models fixed. The distinction is important, as only our approach allows for causal inferences about how changing the treatment assignment mechanism affects policy outcomes, thus providing insights into the comparative value of different content promotion strategies.

Using this ablation approach, we show that even simple data-driven policy, which heuristically mimics the current practice, outperforms it, despite lacking expert judgment or causal inference and prediction capabilities. We further quantify the relative revenue loss of adopting the outcome prediction approach, and further show that a randomized policy – i.e., the treatment assignment mechanism of an A/B test – would have resulted in a revenue loss orders of magnitude greater than the revenue gains of the optimal policy. This suggests that collecting randomized experimental data – which is the standard approach for causal inference and off-policy learning in the marketing literature – may not be economically practical in certain business applications, and this further demonstrates the importance of not only accounting for statistical optimality criteria in choosing between empirical strategies, but also their economic implications.

Our work offers managerial implications for content promotions. Content publishers and producers seeking to optimize their audience-wide content promotions should not necessarily promote the content that is likely to be successful given promotion but, instead, promote the content whose success is likely to *increase the most* if promoted. Our results further suggest how content producers and publishers can improve their audience-wide promotion strategies. For example, a common phrase from the newsroom is “*If it bleeds, it leads*”, implying that news about negative events tends to generate more clicks. Large-scale field experiments support this anecdotal evidence: negative words in news headlines increase click-through rates (Robertson et al. 2023). However, our results call for caution when following such common wisdom: We find that positive stories can also be beneficial to promote when success is not only measured via short-term indicators of engagement (e.g., clicks), but also by longer-term and financially important indicators such as subscriptions.

We foresee valuable extensions of our framework to other marketing contexts. An area where neither personalization nor randomization may be feasible is physical retailing. For example, retailers may not be able to or willing to randomize prices and promotions in-store. However, they typically have access to vast historical data on prices, promotions, and their associated outcomes, and can interventionally adjust these at the the product, category, or store level so as to maximize the

aggregate return across users. In such settings, our framework may be used to learn which items to discount where and when for maximizing incremental store traffic, profits, or repeat purchases, or some other reward function of business importance.

6. Concluding Remarks

In this work, we propose an off-policy learning framework to optimize audience-wide content promotions across digital distribution channels. We present an empirical strategy that connects a model of historical data with non-parametric identification, and we provide a causal machine learning procedure that is unbiased and multiply robust, thus accounting for the difficulties in learning from non-experimental data. We apply our framework to a real-world decision problem at a leading international newspaper, and find that it offers significant performance and revenue gains over current practice. Overall, our work contributes with a robust, scalable, and cost-effective causal machine learning framework to optimize a practically important but relatively understudied marketing decision problem.

Funding and Competing Interests

Author 1 and Author 2 have no competing interests. Author 3 is employed at the partner company.

References

- Agarwal D, Chatterjee S, Yang Y, Zhang L (2015) Constrained Optimization for Homepage Relevance. *WWW Companion* .
- Agarwal D, Chen BC, Elango P (2009) Explore/Exploit Schemes for Web Content Optimization. *IEEE International Conference on Data Mining (ICDM)* .
- Ansari A, Li Y, Zhang JZ (2018) Probabilistic Topic Model for Hybrid Recommender Systems: A Stochastic Variational Bayesian Approach. *Marketing Science* 37(6):987–1008.
- Archak N, Ghose A, Ipeirotis PG (2011) Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science* 57(8):1485–1509.
- Ascarza E (2018) Retention Futility: Targeting High-Risk Customers Might Be Ineffective. *Journal of Marketing Research* 55(1):80–98.
- Athey S, Imbens GW (2016) Recursive Partitioning for Heterogeneous Causal Effects. *Proceedings of the National Academy of Sciences* 113(27):7353–7360.
- Athey S, Tibshirani J, Wager S (2019) Generalized Random Forests. *Annals of Statistics* 47(2):1179–1203.
- Athey S, Wager S (2021) Policy Learning with Observational Data. *Econometrica* 89(1):133–161.
- Battocchi K, Dillon E, Hei M, Lewis G, Oka P, Oprescu M, Syrgkanis V (2019) EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation. Version 0.x.
- Belloni A, Chernozhukov V (2013) Least Squares after Model Selection in High-Dimensional Sparse Models. *Bernoulli* 19(2):521–547.
- Belloni A, Chernozhukov V, Hansen C (2014) Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies* 81(2):608–650.
- Berger J, Humphreys A, Ludwig S, Moe WW, Netzer O, Schweidel DA (2020) Uniting the Tribes: Using Text for Marketing Insight. *Journal of Marketing* 84(1):1–25.

- Berger J, Milkman KL (2012) What Makes Online Content Viral? *Journal of Marketing Research* 49(2):192–205.
- Besbes O, Gur Y, Zeevi A (2016) Optimization in Online Content Recommendation Services: Beyond Click-Through Rates. *Manufacturing and Service Operations Management* 18(1):15–33.
- Bodapati AV (2008) Recommendation Systems with Purchase Data. *Journal of Marketing Research* 45(1):77–93.
- Breiman L (2001) Random Forests. *Machine Learning* 45(1):5–32.
- Center PR (2021) Share of newspaper advertising revenue coming from digital advertising. URL <https://www.pewresearch.org/journalism/chart/sotnm-newspapers-percentage-of-newspaper-advertising-revenue-coming-from-digital/>.
- Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J (2018a) Double/Debiased Machine Learning for Treatment and Structural Parameters. *Econometrics Journal* 21(1):C1–C68.
- Chernozhukov V, Demirer M, Duflo E, Fernández-Val I (2018b) Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, With an Application To Immunization in India.
- DigiDay (2018) How Swiss news publisher NZZ built a flexible pay-wall using machine learning. URL <https://digiday.com/media/swiss-news-publisher-nzz-built-flexible-paywall-using-machine-learning/>.
- DigiDay (2023) Less than half of The Independent’s revenue came from advertising in 2022. URL <https://digiday.com/media/less-than-half-of-the-independents-revenue-came-from-advertising-in-2022/>.
- Dimakopoulou M, Vlassis N, Jebara T (2019) Marginal Posterior Sampling for Slate Bandits. *International Joint Conference on Artificial Intelligence*.
- Dudík M, Erhan D, Langford J, Li L (2014) Doubly Robust Policy Evaluation and Optimization. *Statistical Science* 29(4):485–511.
- Ellickson PB, Kar W, Reeder III JC (2023) Estimating Marketing Component Effects: Double Machine Learning from Targeted Digital Promotions. *Marketing Science* 42(4):704–728.
- EMarketer (2020a) Marketers Expect Content-Driven Campaigns to Increase in 2020. URL <https://www.emarketer.com/content/marketers-expect-content-driven-campaigns-to-increase-in-2020>.
- EMarketer (2020b) US Content Marketing Spending, by Channel, 2019 & 2020. URL <https://www.emarketer.com/chart/241366/us-content-marketing-spending-by-channel-2019-2020-billions-of-total>.
- Foster D, Agarwal A, Dudík M, Luo H, Schapire R (2018) Practical Contextual Bandits with Regression Oracles. *International Conference on Machine Learning*, 1539–1548 (PMLR).
- Foster DJ, Syrgkanis V (2019) Statistical Learning with a Nuisance Component. *Conference on Learning Theory*.
- Friedman JH (2001) Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics* 29(5):1189–1232.
- Garcin F, Dimitrakakis C, Faltings B (2013) Personalized News Recommendation with Context Trees. *ACM Conference on Recommender Systems* 105–112.
- Guhr O (2022) German Sentiment Classification with BERT. URL <https://github.com/oliverguhr/german-sentiment-lib>.
- Guhr O, Schumann AK, Bahrmann F, Böhme HJ (2020) Training a Broad-Coverage German Sentiment Classification Model for Dialog Systems. *Language Resources and Evaluation Conference*.
- Hauser JR, Urban GL, Liberali G, Braun M (2009) Website Morphing. *Marketing Science* 28(2):202–223.
- Hernán M, Robins JM (2020) *Causal Inference: What If* (Boca Raton: Chapman & Hall/CRC).
- Hitsch GJ, Misra S, Zhang WW (2024) Heterogeneous Treatment Effects and Optimal Targeting Policy Evaluation. *Quantitative Marketing and Economics* 22(2):115–168.
- Huang TW, Ascarza E (2024) Doing More with Less: Overcoming Ineffective Long-term Targeting Using Short-Term Signals. *Marketing Science*.

-
- Imbens GW, Rubin DB (2015) *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction* (Cambridge University Press, New York, NY).
- Johar M, Mookerjee V, Sarkar S (2014) Selling vs. Profiling: Optimizing the Offer Set in Web-Based Personalization. *Information Systems Research* 25(2):285–306.
- Kadar C (2022) Fully Redesigned, Algorithmic-driven Next Reads Section. URL <https://medium.com/nzz-open/fully-redesigned-algorithmic-driven-next-reads-section-on-nzz-ch-4501e5919d66>.
- Kale S, Reyzin L, Schapirey RE (2010) Non-Stochastic Bandit Slate Problems. *Advances in Neural Information Processing Systems*.
- Kallus N, Udell M (2020) Dynamic Assortment Personalization in High Dimensions. *Operations Research* 68(4):1020–1037.
- Kennedy EH (2023) Towards Optimal Doubly Robust Estimation of Heterogeneous Causal Effects. *Electronic Journal of Statistics* 17(2):3008–3049.
- Kenton JDMWC, Toutanova LK (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT)*.
- Lada A, Peysakhovich A, Aparicio D, Bailey M (2019) Observational Data for Heterogeneous Treatment Effects with Application to Recommender Systems. *ACM Conference on Economics and Computation*.
- Lemmens A, Gupta S (2020) Managing Churn to Maximize Profits. *Marketing Science* 39(5):956–973.
- Li L, Chu W, Langford J, Schapire RE (2010) A Contextual-Bandit Approach to Personalized News Article Recommendation. *International Conference on World Wide Web*.
- Liu X (2022) Dynamic Coupon Targeting Using Batch Deep Reinforcement Learning: An Application to Livestream Shopping. *Marketing Science* 42(4):637–658.
- Lundberg SM, Lee SI (2017) A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*.
- Murphy SA (2005) A Generalization Error for Q-Learning. *Journal of Machine Learning Research* 6:1073–1097.
- Murphy SA, Van der Laan MJ, Robins JM, Bierman KL, Coie JD, Greenberg MT, Lochman JE, McMahon RJ, Pinderhughes E (2001) Marginal Mean Models for Dynamic Regimes. *Journal of the American Statistical Association* 96(456):1410–1423.
- Oprescu M, Syrgkanis V, Wu ZS (2019) Orthogonal Random Forest for Causal Inference. *International Conference on Machine Learning*.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. (2011) scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.
- Pelger M, Zou J (2022) Inference for Large Panel Data with Many Covariates. *arXiv preprint arXiv:2301.00292*.
- Rafieian O (2022) Optimizing User Engagement Through Adaptive Ad Sequencing. *Marketing Science* 42(5):910–933.
- Rafieian O, Kapoor A, Sharma A (2023) Multi-Objective Personalization of the Length and Skippability of Video Advertisements. *Available at SSRN 4394969*.
- Robertson CE, Pröllochs N, Schwarzenegger K, Parnamets P, Van Bavel JJ, Feuerriegel S (2023) Negativity Drives Online News Consumption. *Nature Human Behaviour* 7(5):812–822.
- Robins JM (1999) Robust Estimation in Sequentially Ignorable Missing Data and Causal Inference Models. *American Statistical Association Section on Bayesian Statistical Science*.
- Robins JM, Rotnitzky A (1995) Semiparametric Efficiency in Multivariate Regression Models with Missing Data. *Journal of the American Statistical Association* 90(429):122.
- Robins JM, Rotnitzky A, Zhao LP (1994) Estimation of Regression Coefficients When Some Regressors Are Not Always Observed. *Journal of the American Statistical Association* 89(427):846–866.

- Rockwell N (2019) News in the Age of Algorithmic Recommendation. URL <https://www.datacouncil.ai/talks/news-in-the-age-of-algorithmic-recommendation>.
- Schwartz EM, Bradlow ET, Fader PS (2017) Customer Acquisition via Display Advertising Using Multi-Armed Bandit Experiments. *Marketing Science* 36(4):500–522.
- Semenova V, Goldman M, Chernozhukov V, Taddy M (2023) Inference on Heterogeneous Treatment Effects in High-Dimensional Dynamic Panels under Weak Dependence. *Quantitative Economics* 14(2):471–510.
- Simester D, Timoshenko A, Zoumpoulis SI (2020a) Efficiently Evaluating Targeting Policies: Improving on Champion vs. Challenger Experiments. *Management Science* 66(8):3412–3424.
- Simester D, Timoshenko A, Zoumpoulis SI (2020b) Targeting prospective customers: Robustness of machine-learning methods to typical data challenges. *Management Science* 66(6):2495–2522.
- Smith AN, Seiler S, Aggarwal I (2022) Optimal Price Targeting. *Marketing Science* 42(3):476–499.
- Song Y, Sahoo N, Ofek E (2019) When and How to Diversify A Multi-Category Utility Model of Consumer Response to Content Recommendations. *Management Science* 65(8):3737–3757.
- Statista (2022) Umsatz der NZZ-Mediengruppe von 2011 bis 2022. URL <https://de.statista.com/statistik/daten/studie/413228/umfrage/umsatz-der-nzz-mediengruppe/>.
- Su F, Mou W, Ding P, Wainwright M (2023) When is the Estimated Propensity Score Better? High-Dimensional Analysis and Bias Correction. *arXiv preprint arXiv:2303.17102*.
- Sutton RS, Barto AG (2018) *Reinforcement Learning: An Introduction* (Cambridge, Massachusetts: MIT press), 2 edition.
- Upworthy (2012) How To Make That One Thing Go Viral. URL [https://www.slideshare.net/Upworthy/how-to-make-that-one-thing-go-viral-just-kidding/25\(2012\)](https://www.slideshare.net/Upworthy/how-to-make-that-one-thing-go-viral-just-kidding/25(2012)).
- Urban GL, Liberali GG, MacDonald E, Bordley R, Hauser JR (2014) Morphing Banner Advertising. *Marketing Science* 33(1):27–46.
- van der Laan MJ, Rose S (2011) *Targeted Learning: Causal Inference for Observational and Experimental Data*, volume 4 (New York: Springer).
- Wager S, Athey S (2018) Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association* 113(523):1228–1242.
- Wang W, Li B, Luo X, Wang X (2023) Deep Reinforcement Learning for Sequential Targeting. *Management Science* 69(9):5439–5460.
- Wasserman L, Roeder K (2009) High Dimensional Variable Selection. *Annals of Statistics* 37(5A):2178–2201.
- Yang J, Eckles D, Dhillon P, Aral S (2024) Targeting for Long-Term Outcomes. *Management Science* 70(6):3841–3855.
- Yoganarasimhan H, Barzegary E, Pani A (2023) Design and Evaluation of Optimal Free Trials. *Management Science* 69(6).
- Zhang Y, Li B, Luo X, Wang X (2019) Personalized Mobile Targeting with User Engagement Stages: Combining a Structural Hidden Markov Model and Field Experiment. *Information Systems Research* 30(3):787–804.
- Zhao S, Witten D, Shojaie A (2021) In Defense of the Indefensible: A Very Naive Approach to High-Dimensional Inference. *Statistical Science* 36(4):562–577.

Online Appendix

Appendix A: Front pages

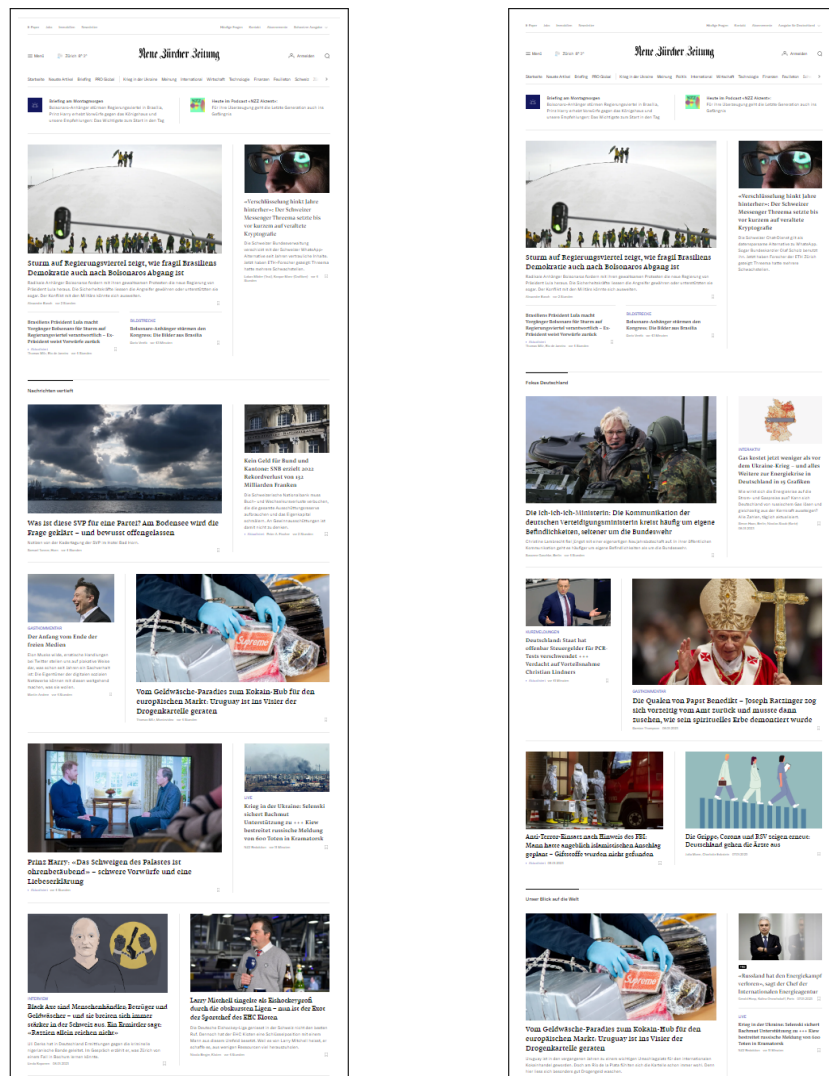


Figure 12 Screenshots of the front page of *Neue Zürcher Zeitung* (due to space, only a few promoted items at the top are shown while the actual list of promoted articles is much longer). Shown: Swiss front page (left) and German front page (right).

Appendix B: Proofs

B.1. Proof of Proposition 1

We now show that Eq. (12) can be written as a function of the observed data distribution. To do so, we use that the density (or, equivalently, the likelihood of a data realization) can be written as a function of both the behavior policy and an arbitrary counterfactual policy.

Let P_π be the joint distribution of the historical data under the behavior policies π_m , $m = 1, \dots, M$. Under our structural model of the DGP, the law of total probability gives that the density of the data in a time period factorizes as

$$p(\mathbf{X}_t, \mathbf{A}_t, Y_t; \pi_1, \dots, \pi_M) = \frac{dP_\pi}{d\mathbb{P}} = p(\mathbf{X}_t) \prod_{m=1}^M \pi_m(A_{tm} | \mathbf{X}_t) p(Y_t | \mathbf{X}_t, \mathbf{A}_t), \quad (33)$$

where we used that a probability density function can be defined as the Radon-Nikodým derivative, $dP_\pi/d\mathbb{P}$, of its cumulative distribution function with respect to a reference measure \mathbb{P} . This definition requires that P_π is absolutely continuous with respect to \mathbb{P} , which holds under Assumption 1.4 (Murphy et al. 2001). Similarly, the density of the data under the joint distribution P_{d_m} induced by a counterfactual policy d_m is given by

$$p(\mathbf{X}_t, \mathbf{A}_t, Y_t; d_m, \pi_k: k \neq m) = \frac{dP_{d_m}}{d\mathbb{P}} = p(\mathbf{X}_t) \mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} \prod_{k \neq m} \pi_k(A_{tk} | \mathbf{X}_t) p(Y_t | \mathbf{X}_t, \mathbf{A}_t), \quad (34)$$

where we again used that P_{d_m} is absolutely continuous with respect to \mathbb{P} . This may be interpreted as that any sequence of realizations that could occur under the counterfactual policy d_m could also have been observed in the data under the behavior policy π_m (Murphy et al. 2001). A change of measure now gives that Eq. (8) can be written as

$$V(d_m) = \frac{1}{T} \sum_{t=1}^T \int \left(y(\mathbf{X}_t, A_{tm}, \mathbf{A}_t^{-m}) \frac{dP_{d_m}}{dP_\pi} \right) dP_\pi = \frac{1}{T} \sum_{t=1}^T \int y_t \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} dP_\pi, \quad (35)$$

where the last equality follows from that

$$\frac{dP_{d_m}}{dP_\pi} = \frac{dP_{d_m}/d\mathbb{P}}{dP_\pi/d\mathbb{P}} = \frac{p(\mathbf{X}_t) \mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} \prod_{k \neq m} \pi_k(A_{tk} | \mathbf{X}_t) p(y_t | \mathbf{X}_t, \mathbf{A}_t)}{p(\mathbf{X}_t) \prod_{m=1}^M \pi_m(A_{tm} | \mathbf{X}_t) p(y_t | \mathbf{X}_t, \mathbf{A}_t)} = \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)}. \quad (36)$$

This is the likelihood ratio (also called the importance weight) between Eq. (34) and Eq. (33), where we used Assumption 1.1 to replace the potential reward with the observed outcome. Eq. (35) defines the value of the counterfactual policy d_m as an expectation with respect to the observed distribution P_π . This shows that the value of an arbitrary counterfactual policy satisfying the assumptions is identified from historical data. Moreover, the identification is non-parametric, as no parametric assumptions have been required to arrive at Eq. (35). What remains to show is that the value function can be written in the AIPW form of Eq. (12). For this, we let

$$\mu(\mathbf{X}_t, d(\mathbf{X}_t), \mathbf{A}_t^{-m}) := \mathbb{E}[Y(\mathbf{X}_t, d(\mathbf{X}_t), \mathbf{A}_t^{-m}) | \mathbf{X}_t] = \mathbb{E}[Y_t | \mathbf{X}_t, d(\mathbf{X}_t), \mathbf{A}_t^{-m}] = \int y_t dP_{Y_t | \mathbf{X}_t, d(\mathbf{X}_t), \mathbf{A}_t^{-m}}(y_t), \quad (37)$$

where the third equality follows from Assumption 1.1. We then note that

$$\mu(\mathbf{X}_t, d_m(\mathbf{X}_t), \mathbf{A}_t^{-m}) = \mathbb{E} \left[Y_t \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} | \mathbf{X}_t \right] \quad (38)$$

$$= \mathbb{E} \left[Y(\mathbf{X}_t, A_{tm}, \mathbf{A}_t^{-m}) \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} \mid \mathbf{X}_t \right] \quad (39)$$

$$= \mathbb{E} \left[Y(\mathbf{X}_t, A_{tm}, \mathbf{A}_t^{-m}) \mid \mathbf{X}_t \right] \mathbb{E} \left[\frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} \mid \mathbf{X}_t \right] \quad (40)$$

$$= \mu(\mathbf{X}_t, A_{tm}, \mathbf{A}_t^{-m}) \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)}, \quad (41)$$

where the first equality holds under Assumption 1.4, the second under Assumption 1.1, and the third by Assumption 1.3, and the last follows by definition of Eq. (37) and basic rules of probability theory. Thus,

$$V(d_m) = \frac{1}{T} \sum_{t=1}^T \int y_t \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} dP_\pi \quad (42)$$

$$= \frac{1}{T} \sum_{t=1}^T \int \left(y_t \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} + \underbrace{\mu(\mathbf{X}_t, d_m(\mathbf{X}_t), \mathbf{A}_t^{-m}) - \mu(\mathbf{X}_t, d_m(\mathbf{X}_t), \mathbf{A}_t^{-m})}_{=0} \right) dP_\pi \quad (43)$$

$$= \frac{1}{T} \sum_{t=1}^T \int \left(y_t \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} + \mu(\mathbf{X}_t, d_m(\mathbf{X}_t), \mathbf{A}_t^{-m}) - \mu(\mathbf{X}_t, A_{tm}, \mathbf{A}_t^{-m}) \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} \right) dP_\pi \quad (44)$$

$$= \frac{1}{T} \sum_{t=1}^T \int \mu(\mathbf{X}_t, d_m(\mathbf{X}_t), \mathbf{A}_t^{-m}) + \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} (y_t - \mu(\mathbf{X}_t, A_{tm}, \mathbf{A}_t^{-m})) dP_\pi, \quad (45)$$

where the third equality follows from our derivations in Eq. (38)–(41). We now see that Eq. (45) is the oracle AIPW formula in Proposition 1. This concludes the proof. \square

B.2. Proof of Theorem 1

We now show that the policy in Eq. (11) is a solution to Eq. (9) and is therefore optimal.

Due to the stationarity of the policies and the myopic nature of the decision problem, the long-run average value is maximized by maximizing the value per time period. That is, for all $t \in \mathcal{V}$,

$$d_m^* = \arg \max_{d_m \in \mathcal{D}} V_t(d_m) \quad (46a)$$

$$\text{s.t. } \sum_{i \in \mathcal{I}_t} d_m(\mathbf{x}_{it}) \leq C_{tm} \quad (46b)$$

with $V_t(d_m)$ given by Eq. (7). Thus, if we can show that the solution to Eq. (46) equals Eq. (11), we are done.

Since $d_m(\mathbf{X}_{it}) \in \{0, 1\}$, we can always decompose the expected reward of an item under a policy decision given a context as the conditional expectation of potential reward if the item would not be promoted plus the CATE if the policy decision is to promote the item, i.e.,

$$\mathbb{E}[Y(\mathbf{X}_{it}, d_m(\mathbf{X}_{it}), \mathbf{A}_{it}^{-m}) \mid \mathbf{X}_{it}] = \mathbb{E}[Y(\mathbf{X}_{it}, A_{itm} = 0, \mathbf{A}_{it}^{-m}) + d_m(\mathbf{X}_{it}) \cdot \tau_m(\mathbf{X}_{it}) \mid \mathbf{X}_{it}], \quad (47)$$

where $\tau_m(\mathbf{X}_{it})$ is given by Eq. (10). The value function for any time period t can thus be written as

$$V_t(d_m) = \mathbb{E}[Y(\mathbf{X}_{it}, A_{itm} = 0, \mathbf{A}_{it}^{-m}) + d_m(\mathbf{X}_{it}) \cdot \tau_m(\mathbf{X}_{it})], \quad (48)$$

where the expectation is over the states. Given the capacity constraint in Eq. (46b), it follows that

$$\max V_t(d_m) = \mathbb{E}[Y(\mathbf{X}_{it}, A_{itm} = 0, \mathbf{A}_{it}^{-m})] + \mathbb{1}\{\tau_m(\mathbf{X}_{it}) \geq \tau_{tm}^{(C_{tm})}\} \cdot \tau_m(\mathbf{X}_{it}). \quad (49)$$

This directly implies that

$$\arg \max_{d_m \in \mathcal{D}} V_t(d_m) = \mathbb{1}\{\tau_m(\mathbf{X}_{it}) \geq \tau_{tm}^{(C_{tm})}\}, \quad (50)$$

which equals d_m^* in Eq. (11) in Theorem 1. This concludes the proof. \square

B.3. Proof of Theorem 2

We now show that the AIPW estimator of policy value given by Eq. (13) is doubly robust. We prove this for an arbitrary policy d_m . Note that, because our procedure for estimating an optimal policy is also unbiased (cf. Proposition 2), the proof implies that the estimated policy value estimate of the optimal policy is doubly robust unbiased of the true value of the true optimal policy, i.e., $V(\hat{d}_m^*) = V(d_m^*)$. To simplify notation, we remove the arguments for the context and remaining channels such that $\mu(a_m) = \mu(\mathbf{X}_t, a_m, A_t^{-m})$, $Y_t(a_m) = Y_t(\mathbf{X}_t, a_m, A_t^{-m})$ for $A = \{A_{tm}, d_m(\mathbf{X}_t)\}$.

First, we first rewrite the expression for the AIPW estimator in Eq. (13):

$$\hat{V}(d_m) = \frac{1}{T} \sum_{t=1}^T \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} \left[\mu(d_m(\mathbf{X}_t)) + \left(Y_t - \mu(A_{tm}) \right) \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} \right] \quad (51)$$

$$= \frac{1}{T} \sum_{t=1}^T \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} \left[\mu(d_m(\mathbf{X}_t)) + Y_t \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} - \mu(A_{tm}) \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} \right] \quad (52)$$

$$= \frac{1}{T} \sum_{t=1}^T \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} \left[\mu(d_m(\mathbf{X}_t)) + Y_t(d_m(\mathbf{X}_t)) \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} - \mu(d_m(\mathbf{X}_t)) \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} \right] \quad (53)$$

$$= \frac{1}{T} \sum_{t=1}^T \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} \left[\underbrace{\mu(d_m(\mathbf{X}_t)) \frac{\pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)}}_{=\mu(d_m(\mathbf{X}_t))} + Y_t(d_m(\mathbf{X}_t)) \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} \right. \\ \left. - \mu(d_m(\mathbf{X}_t)) \frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\}}{\pi_m(A_{tm} | \mathbf{X}_t)} + \underbrace{Y_t(d_m(\mathbf{X}_t)) - Y_t(d_m(\mathbf{X}_t)) \frac{\pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)}}_{=0} \right] \quad (54)$$

$$= \frac{1}{T} \sum_{t=1}^T \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} \left[Y_t(d_m(\mathbf{X}_t)) + Y_t(d_m(\mathbf{X}_t)) \left(\frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} - \pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)} \right) \right] \quad (55)$$

$$- \mu(d_m(\mathbf{X}_t)) \left(\frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} - \pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)} \right) \quad (56)$$

$$= \frac{1}{T} \sum_{t=1}^T \left\{ \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} Y_t(d_m(\mathbf{X}_t)) + \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} \left[\left(Y_t(d_m(\mathbf{X}_t)) - \mu(\mathbf{X}_t, d_m(\mathbf{X}_t)) \right) \right. \right. \quad (57)$$

$$\left. \left(\frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} - \pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)} \right) \right] \right\}, \quad (58)$$

where division by the propensity score is allowed under Assumption 1.4. The third equality follows from that, if $A_{tm} = d_m(\mathbf{X}_t)$, then $Y_t = Y_t(A_{tm}) = Y_t(d_m(\mathbf{X}_t))$ and $\mu(A_{tm}) = \mu(d_m(\mathbf{X}_t))$ by Assumption 1.1. The last equality follows by linearity of expectations and the rest by algebra. We now note that the estimator is unbiased if the expectation of the second term in Eq. (58) is zero, i.e.,

$$\mathbb{E}[\hat{V}(d_m)] = \mathbb{E} \left[\frac{1}{T} \sum_{t=1}^T \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} Y(d_m(\mathbf{X}_{itm})) \right] \quad (60)$$

$$= \mathbb{E}_{P_{d_m}} \left[\frac{1}{T} \sum_{t=1}^T \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} Y(d_m(\mathbf{X}_{itm})) \right] \quad (61)$$

$$= \frac{1}{T} \sum_{t=1}^T \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} Y(d_m(\mathbf{X}_{itm})) \quad (62)$$

$$= V(d_m). \quad (63)$$

Here, the second equality follows from that the expected potential reward given the policy decision equals the expectation of the reward with respect to the distribution under the policy, the third from that the sample average is a constant, and the fourth from the definition of the value.

Now, double robustness means that the estimator is unbiased if (1) $\mu(d_m(\mathbf{X}_t))$ is correctly specified but $\pi_m(A_{tm} | \mathbf{X}_t)$ is not, or (2) $\pi_m(A_{tm} | \mathbf{X}_t)$ is correctly specified but $\mu(d_m(\mathbf{X}_t))$ is not. We have shown that the first term in Eq. (58) is by itself an unbiased estimator. In the following, we thus prove doubly robust unbiasedness by showing that, in either of these cases, the expectation of the second term in Eq. (58) is zero.

Case (1): Assume that $\mu(d_m(\mathbf{X}_t))$ is correctly specified but $\pi_m(A_{tm} | \mathbf{X}_t)$ is not. We then have

$$\mathbb{E} \left[\left(Y_t(d_m(\mathbf{X}_t)) - \mu(d_m(\mathbf{X}_t)) \right) \left(\frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} - \pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)} \right) \right] \quad (64)$$

$$= \mathbb{E} \left[\left(Y_t(d_m(\mathbf{X}_t)) - \mathbb{E}[Y_t | \mathbf{X}_t, d_m(\mathbf{X}_t)] \right) \left(\frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} - \pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)} \right) \right] \quad (65)$$

$$= \mathbb{E} \left[\mathbb{E} \left(\left(Y_t(d_m(\mathbf{X}_t)) - \mathbb{E}[Y_t | \mathbf{X}_t, d_m(\mathbf{X}_t)] \right) \left(\frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} - \pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)} \right) \middle| \mathbf{X}_t, A_{tm} \right) \right] \quad (66)$$

$$= \mathbb{E} \left[\mathbb{E} \left[Y_t(d_m(\mathbf{X}_t)) - \mathbb{E}[Y_t | \mathbf{X}_t, d_m(\mathbf{X}_t)] \middle| \mathbf{X}_t, A_{tm} \right] \left(\frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} - \pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)} \right) \right] \quad (67)$$

$$= \mathbb{E} \left[\left(\mathbb{E}[Y_t(d_m(\mathbf{X}_t)) | \mathbf{X}_t, A_{tm}] - \mathbb{E}[Y_t | \mathbf{X}_t, d_m(\mathbf{X}_t)] \right) \left(\frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} - \pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)} \right) \right] \quad (68)$$

$$= \mathbb{E} \left[\left(\mathbb{E}[Y_t(d_m(\mathbf{X}_t)) | \mathbf{X}_t] - \mathbb{E}[Y_t(d_m(\mathbf{X}_t)) | \mathbf{X}_t] \right) \left(\frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} - \pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)} \right) \right] \quad (69)$$

$$= 0. \quad (70)$$

The first equality used that $\mu(d_m(\mathbf{X}_t)) = \mathbb{E}[Y_t | \mathbf{X}_t, d_m(\mathbf{X}_t)]$ when the reward model is correctly specified, the second follows from iterated expectations, the third and fourth from basic probability theory, and the fifth follows from that $\mathbb{E}[Y_t(d_m(\mathbf{X}_t)) | \mathbf{X}_t, \mathbf{A}_t] = \mathbb{E}[Y_t(d_m(\mathbf{X}_t)) | \mathbf{X}_t]$ by the unconfoundedness assumption and $\mathbb{E}[Y_t | \mathbf{X}_t, d_m(\mathbf{X}_t)] = \mathbb{E}[Y_t(d_m(\mathbf{X}_t)) | \mathbf{X}_t]$ by the consistency assumption. Thus, the two inner expectations cancel each other. The outer empirical expectation and the averaging over stages are omitted from the derivation as they have no effect on the result and the order of expectations can be changed without affecting the result.

Case (2): Now, instead, assume that $\pi_m(A_{tm} | \mathbf{X}_t)$ is correctly specified but $\mu(d_m(\mathbf{X}_t))$ is not. Then, we yield

$$\mathbb{E} \left[\left(Y_t(d_m(\mathbf{X}_t)) - \mu(d_m(\mathbf{X}_t)) \right) \left(\frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} - \pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)} \right) \right] \quad (71)$$

$$= \mathbb{E} \left[\mathbb{E} \left(\left(Y_t(d_m(\mathbf{X}_t)) - \mu(d_m(\mathbf{X}_t)) \right) \left(\frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} - \pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)} \right) \middle| Y_t(d_m(\mathbf{X}_t)), \mathbf{X}_t \right) \right] \quad (72)$$

$$= \mathbb{E} \left[\left(Y_t(d_m(\mathbf{X}_t)) - \mu(d_m(\mathbf{X}_t)) \right) \mathbb{E} \left(\frac{\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} - \pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)} \middle| Y_t(d_m(\mathbf{X}_t)), \mathbf{X}_t \right) \right] \quad (73)$$

$$= \mathbb{E} \left[\left(Y_t(d_m(\mathbf{X}_t)) - \mu(d_m(\mathbf{X}_t)) \right) \frac{\mathbb{E}[\mathbb{1}\{A_{tm} = d_m(\mathbf{X}_t)\} | Y_t(d_m(\mathbf{X}_t)), \mathbf{X}_t] - \pi_m(A_{tm} | \mathbf{X}_t)}{\pi_m(A_{tm} | \mathbf{X}_t)} \right] \quad (74)$$

$$= \mathbb{E} \left[\left(Y_t(d_m(\mathbf{X}_t)) - \mu(d_m(\mathbf{X}_t)) \right) \frac{\{\mathbb{E}[\mathbb{1}\{A_{mt} = d_m(\mathbf{X}_t)\} | \mathbf{X}_t] - \pi_m(A_{mt} | \mathbf{X}_t)\}}{\pi_m(A_{mt} | \mathbf{X}_t)} \right] \quad (75)$$

$$= \mathbb{E} \left[\left(Y_t(d_m(\mathbf{X}_t)) - \mu(d_m(\mathbf{X}_t)) \right) \frac{\{\pi_m(A_{mt} | \mathbf{X}_t) - \pi_m(A_{mt} | \mathbf{X}_t)\}}{\pi_m(A_{mt} | \mathbf{X}_t)} \right] \quad (76)$$

$$= 0, \quad (77)$$

where the first equality follows by iterated expectations, the second and third by basic probability theory, the fourth by Assumption 1.2, and the fifth by the correct specification of each propensity score model so that the terms in the numerator cancel. This concludes the proof. \square

Appendix C: Descriptive Statistics

Tables 5–6 show summary statistics of all variables, grouped by the training set and test set. Recall that non-categorical variables are scaled to zero mean and unit variance. The high outliers are from articles on COVID-19, which received above attention traction among the user base.

Covariate	Training set				Test set			
	Mean	Std	Min	Max	Mean	Std	Min	Max
Article views	200.07	525.66	0.00	50124.00	211.43	485.98	0.00	21716.00
Average engagement time	1.04	1.90	0.00	374.80	0.94	0.83	0.00	48.64
Average scroll depth	63.87	26.81	0.00	100.00	63.70	26.92	0.00	100.00
Recirculation	30.96	22.60	0.00	100.00	30.67	22.61	0.00	100.00
Cumulative registrations	0.32	1.27	0.00	44.00	0.11	0.56	0.00	29.00
Number of times on the subscription path	0.01	0.06	0.00	1.80	0.01	0.05	0.00	2.20
Content length	1145.82	1704.14	0.00	41173.00	1179.12	1912.07	0.00	41173.00
Hours online on Swiss front page	339.26	4955.98	0.00	182219.00	227.83	1404.67	0.00	15825.00
Hours online on German frontpage	312.37	4944.18	0.00	182220.00	200.04	1357.71	0.00	15836.00
Hours online since last publication	33.12	20.81	-0.52	71.94	32.40	20.74	-0.25	71.92
Count of email newsletter promotions	8.99	52.50	0.00	502.00	9.54	54.20	0.00	515.00
Count of push notification promotions	1.96	14.22	0.00	192.00	2.23	15.71	0.00	202.00
Count of Twitter promotions	4.96	28.69	0.00	333.00	5.13	29.24	0.00	343.00
Swiss front page promotion	0.17	0.37	0.00	1.00	0.17	0.37	0.00	1.00
German front page promotion	0.17	0.37	0.00	1.00	0.17	0.38	0.00	1.00
Lag of Swiss front page promotion	0.17	0.38	0.00	1.00	0.17	0.38	0.00	1.00
Lag of German front page promotion	0.17	0.37	0.00	1.00	0.17	0.38	0.00	1.00
Lag of Email newsletter promotion	0.01	0.09	0.00	1.00	0.01	0.10	0.00	1.00
Lag of Push notification promotion	0.00	0.06	0.00	1.00	0.00	0.06	0.00	1.00
Lag of Twitter promotion	0.01	0.10	0.00	1.00	0.01	0.11	0.00	1.00

Table 5 Summary statistics. Note: non-categorical variables were standardized.

Category	Training	Evaluation
Hour: 06:00	0.05	0.06
Hour: 07:00	0.05	0.06
Hour: 08:00	0.05	0.06
Hour: 09:00	0.06	0.06
Hour: 10:00	0.06	0.06
Hour: 11:00	0.06	0.06
Hour: 12:00	0.06	0.06
Hour: 13:00	0.06	0.06
Hour: 14:00	0.06	0.06
Hour: 15:00	0.06	0.06
Hour: 16:00	0.06	0.05
Hour: 17:00	0.05	0.06
Hour: 18:00	0.06	0.05
Hour: 19:00	0.06	0.06
Hour: 20:00	0.06	0.06
Hour: 21:00	0.06	0.06
Hour: 22:00	0.06	0.05
Hour: 23:00	0.05	0.04
Weekday: Monday	0.10	0.08
Weekday: Tuesday	0.09	0.09
Weekday: Wednesday	0.09	0.14
Weekday: Thursday	0.19	0.24
Weekday: Friday	0.20	0.25
Weekday: Saturday	0.18	0.11
Weekday: Sunday	0.14	0.09

Table 6 Share of items belonging to each categorical covariate of time information

Category	Training	Evaluation
Section: Culture	0.01	0.00
Section: Education	0.09	0.09
Section: Society	0.01	0.01
Section: International	0.22	0.25
Section: Opinion	0.09	0.08
Section: Mobility	0.01	0.01
Section: NZZ in English	0.01	0.01
Section: Other	0.01	0.01
Section: International politics	0.02	0.02
Section: Celebrities & events	0.08	0.08
Section: Domestic	0.08	0.10
Section: Sport	0.08	0.08
Section: Visuals	0.01	0.01
Section: Business & finance	0.15	0.13
Section: Science & technology	0.05	0.04
Section: Zurich	0.08	0.07
Type: None	0.71	0.73
Type: Editor-in-Chief	0.01	0.01
Type: Breaking news	0.08	0.09
Type: English	0.01	0.01
Type: Explained	0.01	0.00
Type: Guest comment	0.03	0.02
Type: Interview	0.03	0.02
Type: Column	0.01	0.01
Type: Comment	0.08	0.06
Type: News in brief	0.03	0.03
Type: Other	0.01	0.01
Format: Long-form standard	0.10	0.08
Format: Long-form visual	0.01	0.01
Format: Opinion	0.11	0.10
Format: Regular	0.78	0.81
Format: Video	0.00	0.00
Sentiment: Negative lead_text	0.08	0.08
Sentiment: Neutral lead_text	0.92	0.92
Sentiment: Positive lead_text	0.00	0.00
Sentiment: Negative SEO title	0.09	0.06
Sentiment: Neutral SEO title	0.89	0.93
Sentiment: Positive SEO title	0.02	0.01
Sentiment: Negative title	0.07	0.05
Sentiment: Neutral title	0.91	0.94
Sentiment: Positive title	0.02	0.01

Table 7 Share of items belonging to each categorical covariate of content characteristics.

Appendix D: Estimation, Selection and Tuning of Machine Learning Models

D.1. Nuisance Models

The nuisance models are not of interest in themselves but simply are means to predict potential outcomes. Hence, we select nuisance models based on predictive performance. As candidate models, we considered random forests (Breiman 2001) and gradient-boosted trees (Friedman 2001). Both are ensembles of non-parametric decision trees, but differ in their construction. Eventually, we found that gradient-boosted trees perform better. Nevertheless, repeating the analyses from our main paper with random forests leads to qualitatively similar findings.

We perform hyperparameter tuning and model selection as follows. First, we perform 5-fold cross-validation based on an exhaustive grid search to find the hyperparameter values that minimize the loss function of each nuisance model on the training data. For the propensity score model, we search for the hyperparameters that minimize the logistic loss, given by

$$\text{Logistic loss} = \frac{1}{C} \sum_{c=1}^C \frac{1}{N_c} \sum_{i \in c} \left(A_{im}^{(c)} \log \pi_{im}^{(c)} + (1 - A_{im}^{(c)}) \log(1 - \pi_{im}^{(c)}) \right), \quad (78)$$

where $\pi_{im}^{(c)} = \pi_m(A_{im}^{(c)} | \mathbf{X}_i^{(c)})$ is short-hand notation for the propensity score of promotion on channel m on validation observation i in cross-validation fold $c = 1 \dots, C$ to channel $m = 1, \dots, M$. For the reward model, we search for the hyperparameter values that minimize the mean squared error (MSE) loss, given by

$$\text{MSE loss} = \frac{1}{C} \sum_{c=1}^C \frac{1}{N_c} \sum_{i \in c} \left(\varepsilon_{im}^{(c)} \right)^2, \quad (79)$$

where $\varepsilon_{im}^{(c)} = Y_i^{(c)} - \mu(\mathbf{X}_i^{(c)}, A_{im}^{(c)}, \mathbf{A}_i^{(c), -m})$ is the error of the reward model. Second, we compare the hyperparameter-tuned nuisance models in terms of their predictive ability. For this, we take the average loss across the validation splits in the cross-validation. Based on this we select one propensity score model and one reward model. Details on the hyperparameter tuning for the propensity score models and the reward models are shown in Table 8 and 9, respectively.

Hyperparameter	Grid-searched values	Tuned hyperparameter values			
		Random Forest		Gradient Boosted Trees	
		Switzerland	Germany	Switzerland	Germany
n_estimators	[100, 200, 300, 400, 500]	500	300	200	500
max_features	[sqrt, log2, None]	None	sqrt	log2	sqrt
max_depth (RF)	[20, 40, 60, 80, 100, None]	20	40		
max_depth (GB)	[3, 5, 7, 9]			9	9
min_samples_split	[2, 4, 8, 16]	2	16	4	2
min_samples_leaf	[1, 2, 4, 8]	8	4	1	8
learning_rate	[0.01, 0.05, 0.1, 0.2]			0.05	0.01

Table 8 Details on hyperparameter tuning for the propensity score models. The hyperparameters were tuned to minimize the loss function via 5-fold cross-validated grid search on the training data. See the documentation for scikit-learn for descriptions of the hyperparameter. Empty cells mean that the hyperparameter does not apply to the model class. Abbreviations: RF = Random Forest. GBT = Gradient Boosted Trees.

Hyperparameter	Grid-searched values	Tuned hyperparameter values	
		Random Forest	Gradient Boosted Trees
<code>n_estimators</code>	[100, 200, 300, 400, 500]	500	300
<code>max_features</code>	[<i>sqrt</i> , <i>log2</i> , <i>None</i>]	None	<i>sqrt</i>
<code>max_depth</code> (RF)	[20, 40, 60, 80, 100, <i>None</i>]	20	
<code>max_depth</code> (GB)	[3, 5, 7, 9]		7
<code>min_samples_split</code>	[2, 4, 8, 16]	2	2
<code>min_samples_leaf</code>	[1, 2, 4, 8]	4	1
<code>learning_rate</code>	[0.01, 0.05, 0.1, 0.2]		0.05

Table 9 Details on hyperparameter tuning for the reward models. The hyperparameters were tuned to minimize the loss function via 5-fold cross-validated grid search on the training data. See the documentation for scikit-learn (Pedregosa et al. 2011) for descriptions of the hyperparameter. Empty cells mean that the hyperparameter does not apply to the model class.

D.2. Hyperparameter tuning of CATE Model

We tune the hyperparameters of the orthogonal random forest CATE function as follows. We set the total number of trees to 500. We leave the maximum depth of trees unrestricted such that nodes are expanded until leaves are pure or until all leaves contain less than `min_samples_split` samples. Then, we fix the minimum number of splitting samples required to split an internal node to two and the minimum number of samples required to be at a leaf node to one. We specify the model to consider all features when looking for the best split. The kernel is tuned as part of the 5-fold cross-validated estimation.

D.3. Best Linear Predictor of CATE with Post-Lasso Inference

We perform the two-stage procedure from the post-Lasso separately for the Swiss and German front page policies using the CATE estimates of either as the dependent variable. We tune the Lasso regularized model using 5-fold cross-validation. For each cross-validation run, we run coordinate descent optimization for 1000 iterations with 500 values of the penalty λ along different regularization values (here, we use the default values from scikit-learn). The final model is selected as the respective model that best predicts the BLP of the CATE among the corresponding 10 cross-validated models. For inference, we compute HC2 robust standard errors to guard against heteroskedasticity in the second-stage error terms and obtain p -values based on those. We implement the method using the python package scikit-learn (Pedregosa et al. 2011).

D.4. Sample-Split Cross-Fitting

We implement a tailored sample-split cross-fitting procedure with the following steps:

1. For cross-validation fold $l = \dots, L$:
 - (a) Randomly split the training data \mathcal{T} into non-overlapping sets $\mathcal{T}_1^{(l)}$ and $\mathcal{T}_2^{(l)}$ of equal size $|\mathcal{T}|/2$ such that $\mathcal{T}_1^{(l)} \cup \mathcal{T}_2^{(l)} = \mathcal{T}$
 - (b) On $\mathcal{T}_1^{(l)}$, fit the hyperparameter-tuned nuisance models in Eqs. (17)–(18)
2. For cross-validation fold $l = \dots, L$:
 - (a) For each item $i \in \mathcal{T}_2^{(l)}$, predict the doubly robust scores by evaluating Eq. (19) for $a \in \{0, 1\}$ using an reward model and a nuisance model that were fitted on different folds $k, j \in \{1, \dots, L\}$, $k \neq j$, not yet paired.
 - (b) Use Eq. (20) to obtain CATE estimates on fold l ;
3. On $\bigcup_{l=1}^L \mathcal{T}_2^{(l)}$, the union of nuisance test sets across folds, fit the CATE function using Eq. (21). We run the procedure using $L = 5$ cross-validation folds.

D.5. Software Implementation

We perform all analyses in Python. We implement the nuisance models using the package scikit-learn (Pedregosa et al. 2011), the orthogonal random forest in our CATE function via the package EconML (Battocchi et al. 2019), and the post-Lasso using the packages scikit-learn and statsmodels.

Appendix E: Additional Figures

Figures 13–17 show additional results for the CATE estimates of the optimal policy and its value, both in aggregate across all content and across the subset of content that is promoted or not promoted by the optimal policy.

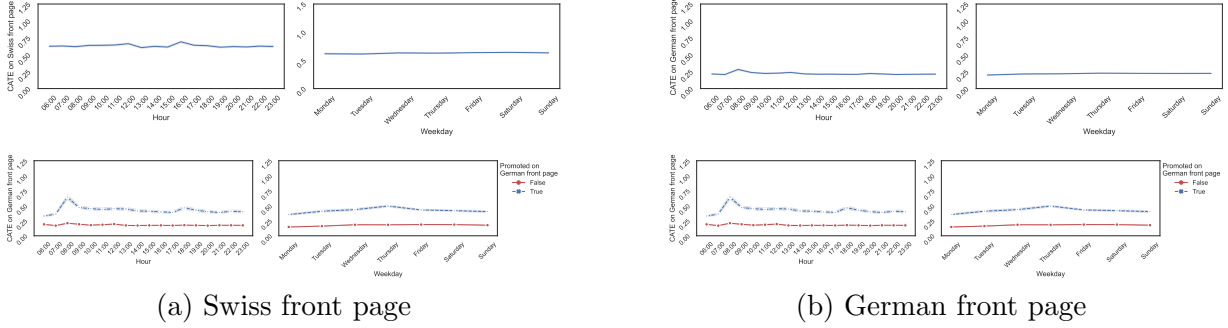


Figure 13 Estimated CATE over time for all content (top) and for content promoted or not promoted by the optimal policy (bottom). Error bars are 95% confidence intervals across 1000 bootstrap runs.

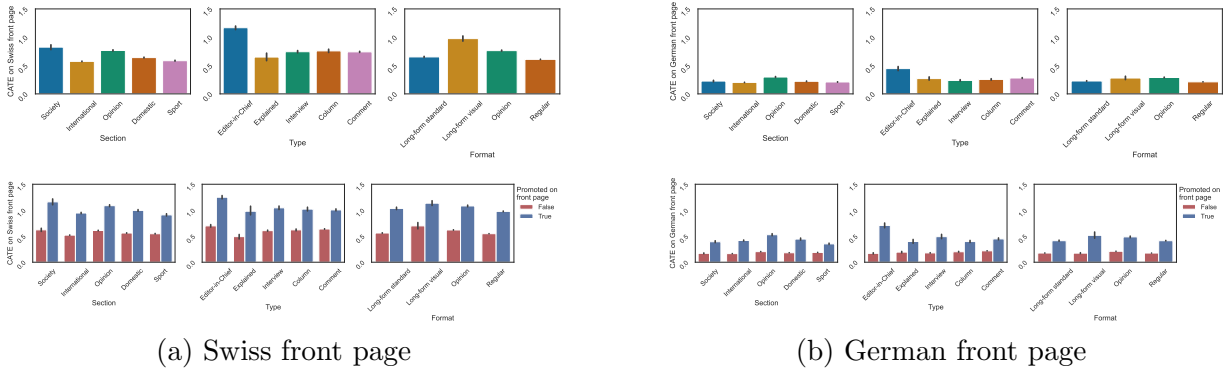


Figure 14 Estimated CATE by content categories for all content and for content promoted and not promoted by the optimal policy. Error bars are 95% confidence intervals across 1000 bootstrap runs.

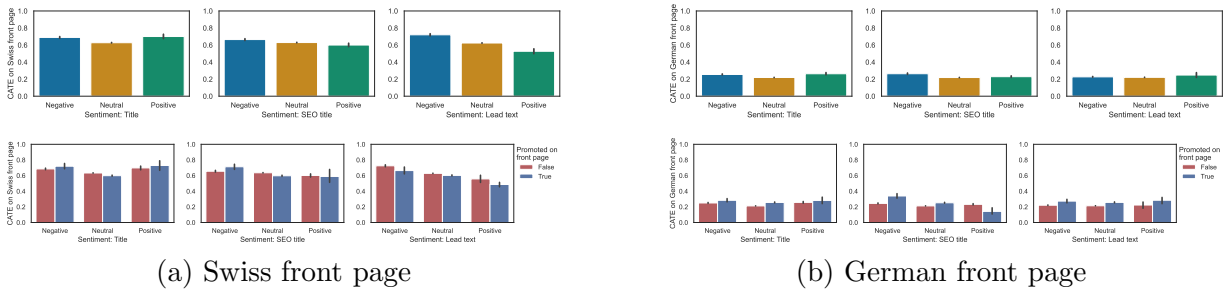
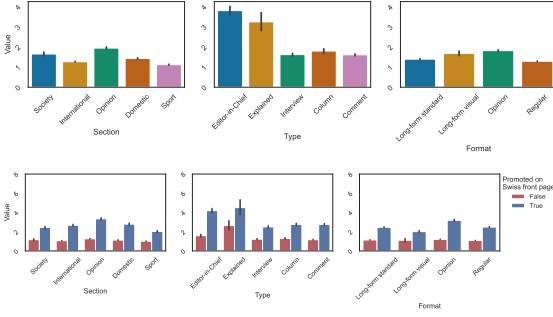
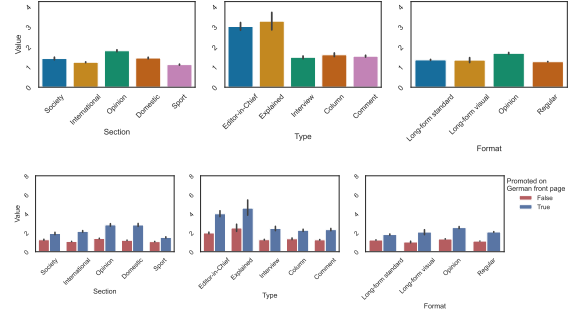


Figure 15 Estimated CATE by content sentiment for all content and for content promoted and not promoted by the optimal policy. Error bars are 95% confidence intervals across 1000 bootstrap runs.

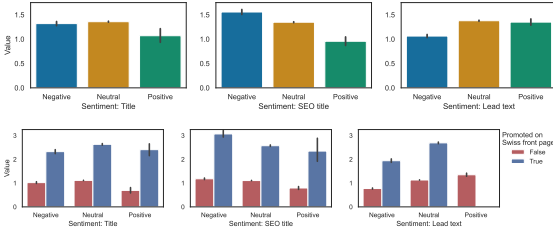


(a) Swiss front page

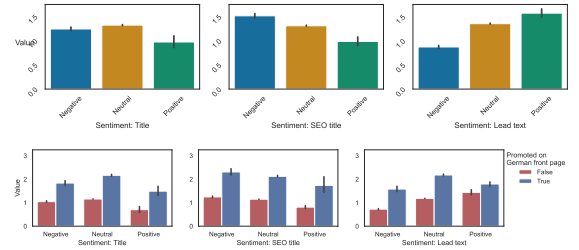


(b) German front page

Figure 16 Value of optimal policy by categories for all content and for content promoted and not promoted by the optimal policy. Error bars are 95% confidence intervals across 1000 bootstrap runs.



(a) Swiss front page



(b) German front page

Figure 17 Value of optimal policy by sentiment for all content and for content promoted and not promoted by the optimal policy. Error bars are 95% confidence intervals across 1000 bootstrap runs.