

---

## Data Mining for Marketing Analytics

Period: a.y. 2023/24 – IV sem.

Class times: 16:30 – 18:00

Monday, Tuesday, Wednesday

**Instructor:**

Prof. Kai Zhu

Dept. of Marketing. - Room 4-D1-15

[kai.zhu@unibocconi.it](mailto:kai.zhu@unibocconi.it)

---

### Course description

Large Language Models have revolutionized the landscape of data mining and Marketing Research. They form the basis of all state-of-the-art systems across a wide range of tasks and have shown an impressive ability to generate fluent text and perform few-shot learning. Not only they are great tools for social analytics, but at the same time those large pre-trained models have created enormous social impacts across a wide range of fields. This course will introduce the fundamentals of large language models, overview their societal impacts, and explore potential research opportunities around AI in our modern society. In this course, students will learn the fundamentals about the theory, data, general capabilities, and social risk aspects of large language models, as well as gain hands-on experience working with them for marketing research.

### Course Material

The content of the sessions and the slides are the only required material. There is also a list of suggested readings that are meant to provide a range of perspectives and flavors of the research of large language models. The slides and the readings will be made available on Blackboard.

### Tentative list of topics

**Lecture 1+2:** Introduction to Neural Networks

**Lecture 3+4:** Word Embedding in Social Science

**Lecture 5+6:** Neural Embedding for Marketing Research

**Lecture 7+8:** Understanding Large Language Model

**Lecture 9+10:** Societal Impact of Large Language Model

**Lecture 11+12:** Ethical Consideration and Social Risk of Harm

### **Assessment Methods.**

Effective class participation includes attendance to the sessions and making an active and constructive contribution to the discussion, asking questions, making constructive comments, and having a positive attitude toward learning. As we go through the course, students will also be required to participate to paper presentation. Students will be evaluated as follows:

Class participation:	20 points
Hands-on exercise:	20 points
Paper discussion:	30 points
Final research proposal:	30 points



## Suggested Readings

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183-186.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16), E3635-E3644.

Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review*, 84(5), 905-949.

Toubia, O., Berger, J., & Eliashberg, J. (2021). How quantifying the shape of stories predicts their success. *Proceedings of the National Academy of Sciences*, 118(26), e2011695118.

Waller, I., & Anderson, A. (2019, May). Generalists and specialists: Using community embeddings to quantify activity diversity in online platforms. In *The World Wide Web Conference* (pp. 1954-1964).

Waller, I., & Anderson, A. (2021). Quantifying social organization and political polarization in online platforms. *Nature*, 600(7888), 264-268.

Timoshenko, A., & Hauser, J. R. (2019). Identifying customer needs from user-generated content. *Marketing Science*, 38(1), 1-20.

McAuley, J., Targett, C., Shi, Q., & van den Hengel, A. (2015). P2V-MAP: Mapping Market Structures for Large Retail Assortments. *Proceedings of the ACM Conference on Electronic Commerce*.

Aral, S., Eckles, D., & Kumar, M. (2019). Scalable bundling via dense product embeddings. MIT Working Paper.

Chen, F., Liu, X., Proserpio, D., & Troncoso, I. (2022). Product2Vec: Leveraging representation learning to model consumer product choice in large assortments. NYU Stern School of Business.

Korinek, A. (2023). Language Models and Cognitive Automation for Economic Research (No. w30957). National Bureau of Economic Research.

Horton, J. J. (2022). Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?.

Kosinski, M. (2023). Theory of Mind May Have Spontaneously Emerged in Large Language Models. arXiv preprint arXiv:2302.02083.

Schramowski, P., Turan, C., Andersen, N., Rothkopf, C. A., & Kersting, K. (2022). Large pre-trained language models contain human-like biases of what is right and wrong to do. *Nature Machine Intelligence*, 4(3), 258-268.

Abid, A., Farooqi, M., & Zou, J. (2021). Large language models associate Muslims with violence. *Nature Machine Intelligence*, 3(6), 461-463.

Cadario, R., Longoni, C., & Morewedge, C. K. (2021). Understanding, explaining, and utilizing medical artificial intelligence. *Nature human behaviour*, 5(12), 1636-1642.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, 46(4), 629-650.

Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, 9(1), 104-117.

McGuffie, K., & Newhouse, A. (2020). The radicalization risks of GPT-3 and advanced neural language models. *arXiv preprint arXiv:2009.06807*.

Meta Fundamental AI Research Diplomacy Team (FAIR)<sup>†</sup>, Bakhtin, A., Brown, N., Dinan, E., Farina, G., Flaherty, C., ... & Zijlstra, M. (2022). Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science*, 378(6624), 1067-1074.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., ... & Gabriel, I. (2021). Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359*.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., ... & Fedus, W. (2022). Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.

Research: quantifying GitHub Copilot's impact on developer productivity and happiness

- Link: <https://github.blog/2022-09-07-research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/>

Noy, S., & Zhang, W. (2023). Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *Available at SSRN 4375283*.