Bocconi University

PhD School

Data Mining for Marketing Analytics

or AI-Assisted Research Methods

Course Code:	Ph.D. Course 41061
Period:	a.y. 2025/26 – II semester
Sessions:	12 sessions
Instructor:	Prof. Kai Zhu
Department:	Marketing, Room 4-D1-09
Email:	kai.zhu@unibocconi.it
Office Hours:	By appointment

1 Course Overview

This doctoral seminar introduces AI-assisted research methods with a focus on large language models (LLMs) and agentic AI systems for quantative research. The course bridges cutting-edge AI capabilities with rigorous academic research practices, teaching PhD students how to leverage tools like Claude Code, ChatGPT, and custom AI agents to enhance every stage of the research process—from literature review and data collection to analysis, writing, and reproducibility.

Unlike introductory AI courses, this seminar emphasizes *research methodology*, *validation*, and *ethical considerations* specific to academic work. Students will learn not only *how* to use AI tools effectively, but also *when* to trust them, how to validate AI-generated outputs, and how to maintain scientific rigor while benefiting from automation and augmentation.

2 Course Philosophy

This course is built on three core principles. First, AI is treated as a form of research augmentation rather than a replacement for human judgment: AI tools are intended to amplify researchers' capabilities, not to substitute for critical thinking, domain expertise, or theoretical insight. Second, the course prioritizes rigor over efficiency. While speed and automation are valuable, they are never pursued at the expense of methodological soundness, reproducibility, or intellectual honesty. Third, the course emphasizes practical application to real research. Every technique taught is designed to directly support students' dissertation work and to remain relevant to their future academic careers.

3 Prerequisites

Students are expected to have completed core PhD coursework in quantitative research methods. They should possess basic familiarity with data analysis, although no advanced programming skills are required. Prior experience writing research papers is assumed. In addition, students should demonstrate an openness to experimentation, including a willingness to engage with new tools and to iteratively refine their research workflows.

4 Course Structure

The course consists of 12 sessions organized into four modules:

Module 1: Foundations of AI-Assisted Research

Module 2: AI for Literature and Theory

Module 3: Module 3: AI for Data & Analysis

Module 4: Writing, Ethics & Research Impact

5 Assessment

Students will be evaluated based on the following components:

Component	Weight	Due Date
Class Participation & Discussion	20%	Ongoing
Hands-On Research Exercises	25%	Throughout semester
Paper Presentations & Leading Discussion	25%	Sessions 3–11
Final Research Project	30%	End of semester

6 Course Policies

Attendance & Participation

Given the seminar format and small cohort size, attendance at all sessions is expected. If you must miss a session due to conference travel or other academic obligations, please notify the instructor in advance and arrange to make up the work.

AI Use Policy

AI use is not just allowed—it is the *entire point* of this course. Students are expected to understand what tools such as Claude produce for them; "the AI did it" is not an acceptable explanation for submitted work. All AI-generated outputs must be validated, and the course will explicitly teach the habits and methods required for proper validation. In addition, AI-generated text may not be presented as a student's own intellectual contribution without meaningful critical engagement, revision, and proper attribution.

Collaboration

Students are encouraged to collaborate on troubleshooting technical issues, discussing readings, and providing feedback on each other's work.

7 Tentative Reading List

Foundational & Methodological

- Ziems, C., et al. (2024). "Can Large Language Models Transform Computational Social Science?" *Computational Linguistics*.
- Bail, C. A. (2024). "Can Generative AI Improve Social Science?" *Proceedings of the National Academy of Sciences*.
- Argyle, L. P., et al. (2023). "Out of One, Many: Using Language Models to Simulate Human Samples." *Political Analysis*.
- Binz, M., & Schulz, E. (2023). "Using Cognitive Psychology to Understand GPT-3." *Proceedings of the National Academy of Sciences*.
- Tornberg, P. (2024). "How to Use LLMs for Text Analysis." *Sociological Methods & Research* (forthcoming).

Marketing Applications

- Garg, N., et al. (2024). "Large Language Models for Marketing Research: Applications, Opportunities, and Challenges." Working paper.
- Matz, S. C., et al. (2024). "The Potential of Generative AI for Personalized Persuasion at Scale." *Scientific Reports*.
- Simester, D., et al. (2020). "How AI Will Change the Future of Marketing." *Harvard Business Review*.
- Huang, M. H., & Rust, R. T. (2021). "A Strategic Framework for Artificial Intelligence in Marketing." *Journal of the Academy of Marketing Science*.
- Davenport, T., et al. (2020). "How Artificial Intelligence Will Change the Future of Marketing." *Journal of the Academy of Marketing Science*.

Text Analysis & NLP

- Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press. (Selected chapters)
- Gilardi, F., et al. (2023). "ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks." *Proceedings of the National Academy of Sciences*.
- Törnberg, P., Valeeva, D., Uitermark, J., & Bail, C. (2024). "Simulating Social Media Using Large Language Models to Evaluate Alternative News Feed Algorithms." arXiv preprint.
- Berger, J., et al. (2020). "Uniting the Tribes: Using Text for Marketing Insight." *Journal of Marketing*.

Experimental Design & Surveys

- Horton, J. J. (2023). "Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?" NBER Working Paper.
- Aher, G. V., Arriaga, R. I., & Kalai, A. T. (2023). "Using Large Language Models to Simulate Multiple Humans and Replicate Human Subject Studies." *ICML*.
- Brand, J. E., et al. (2023). "Can Large Language Models Predict Consumer Behavior?" Working paper.
- Dillion, D., et al. (2023). "Can AI Language Models Replace Human Participants?" *Trends in Cognitive Sciences*.

Ethics, Bias & Validation

- Bender, E. M., et al. (2021). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *FAccT*.
- Bommasani, R., et al. (2021). "On the Opportunities and Risks of Foundation Models." Stanford CRFM Report.
- Gershgorn, D., & Krolik, A. (2024). "The AI Disclosure Dilemma in Academic Publishing." *Science*.
- Narayanan, A., & Kapoor, S. (2023). "GPT-4 and Professional Benchmarks: The Wrong Approach to Measurement." AI Snake Oil blog.
- Liang, W., et al. (2023). "Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on Scientific Publications." arXiv preprint.

Tools & Technical

- Anthropic (2024). "Claude Code Documentation." https://docs.anthropic.com
- White, J., et al. (2023). "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT." arXiv preprint.
- Liu, P., et al. (2023). "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing." *ACM Computing Surveys*.
- OpenAI (2023). "GPT-4 Technical Report." arXiv preprint.

Note: This reading list is tentative and will be updated with the latest research as the field evolves rapidly. Additional readings will be provided on Blackboard.

8 Acknowledgments

This course draws inspiration from several pioneering initiatives in AI for research:

- Christopher Bail's "Large Language Models for Social Science" (Duke University)
- Sendhil Mullainathan's "AI and Economics" seminar (MIT)

- The Computational Social Science community at large

This syllabus is subject to revision. Students will be notified of any changes via email and Blackboard.