Experiences with experimental research Lessons from a co-creative research project on AI effects in companies

March 2025

ai:conomics policy*brief*

Experiences with experimental research Lessons from a co-creative research project on AI effects in companies

Marie-Christine Fregin, Nicholas Rounding, Rolf van der Velden, Mark Levels, Pelin Özgül and Sanne Steens

About the "ai:conomics" research project

The use of intelligent technologies in companies is changing the way we work. Artificial intelligence (AI) has a direct impact on how work is organised. This leads to changes in work tasks, skill requirements and productivity. AI also has an indirect impact on working conditions and the well-being of employees in the workplace. The transdisciplinary research project ai:conomics aims to expand the scientific knowledge about the effects of AI on work and employees and make it available to a broad audience. The project thus creates a better, evidence-based foundation to better shape the future of work and the use of human-centred AI (further information can be found on the project website.

To achieve this goal, ai:conomics conducts controlled field studies in large organisations in Germany and the Netherlands that have implemented AI tools for different purposes. The research project measures the direct impact of AI implementation on employees' job characteristics, skill profiles, preferences, and well-being as well as on productivity using key performance indicators (KPIs). To contextualise the results, the individual findings are related to general developments on the German labour market and supplemented by differentiated analyses of AI-related changes using German register data.

The unique character of the ai:conomics project is defined by the co-creative process that enables the project to be *jointly* designed through transdisciplinary work between researchers, employers, employees, technology experts, works councils and policy makers. Together with many different stakeholders in business and politics, the ai:conomics team aims to gain new insights and share the knowledge gained with all interest groups, experts and the general public. This will ensure that as many people as possible benefit from the empirical findings on the effects of the use of AI in the world of work.

Learning from evaluating AI implementation processes in ai:conomics

The ai:conomics field studies are based on the staggered introduction of AI at different locations of each company. These field studies allow us to identify experimental and control groups and measure changes in the performance, preferences, and well-being of employees in the period before and after the introduction of AI. As AI continues to learn, we cannot expect homogeneous effects over time. Instead, we need to model heterogeneous effects between early and later applications of the AI tools in some studies. The staggered design of AI adoption allows us to measure these heterogeneous "learning effects" by distinguishing between the early and later applications of AI tools within the same organisation.

We started the ai:conomics project at the time when we had signed letters of intent (LOIs) from the companies. After several meetings in which we presented and explained the scientific approach of our project, the companies agreed in writing that experimental field studies would be carried out in their company to accompany and evaluate the AI implementation.

Why are the companies taking part? The most important arguments (Unique Selling Points) of our project for the companies:

- ai:conomics provides sound scientific findings on the impact of AI on employees, work and productivity, which companies can use to make strategic decisions.
- Our project serves as a "neutral ground" for the interests of all relevant stakeholder groups (works councils, management, employees).
- To increase the positive impact of AI, companies have an interest in participating in our project and learning from AI implementation in their own and other companies and using this experience for future AI projects.

Companies, policy makers and researchers have an interest in obtaining *causal* evidence of the impact of AI on employees, labour, and productivity. At the same time, however, the complex environment in organisations makes it difficult to provide this evidence. In this policy brief, we present some of the insights we have gained from the project on co-creative research on the effects of AI. Co-creative research is an innovative evaluation design which may help researchers, policymakers and companies to get clear results. Yet, it also pertains some challenges and this policy brief may help to illustrate some of the different viewpoints between business and academia, the challenges, and limitations of field research in a high-tech environment and some of the steps we have taken to overcome the challenges involved. We also hope that our lessons learnt will be useful for researchers considering such fieldwork. And that our experiences will help to identify future opportunities for randomised experiments within the overall process of introducing and applying new AI tools.

Since the complexity of such field experiments is not unique to the world of AI but can also be found in many other fields in which social scientists conduct research (education, economics, healthcare, administration), we believe that the insights gained are also valuable for social scientists in other fields and disciplines interested in evaluating causal effects of interventions in complex environments. However, it should be recognised that the settings of our study may be very different from those in other sectors, e.g., in terms of the ability to conduct randomised control trials, the availability of production data at the company, group or individual level, or the external validity of experiments (e.g., schools are more similar than companies). There is also a big difference between AI that both increases productivity and changes the employment structure (i.e., the number and characteristics of workers before and after the introduction of AI are different) and interventions that only aim to increase the performance of people already involved before the intervention. The latter in particular were the focus of ai:conomics.

Based on our experience with the case studies in ai:conomics, we have drawn several lessons that can be categorised into four groups or clusters: 1. complexity of AI implementation; 2. complexity of the co-creation process; 3. methodological complexity; and 4. legal and ethical complexities. However, it should be noted that these clusters are closely linked. The complexity of AI implementation necessitates co-creation and has a direct impact on methodological complexity: the different lessons learnt are thus closely intertwined. These lessons illustrate how challenges that we dealt with when setting up experimental studies in different companies. The lessons are not intended to discourage researchers in this field to conduct such experiments. Instead, we would like to emphasize that maybe the most important lesson of all, is that it is really worth to deal with these lessons.

Learning experiences on different levels
Cluster 1: Complexity of AI implementation

> **Lesson 1:** Implementing AI is not a simple, unidirectional and clear-cut process: Expect the unexpected and adapt the research strategy flexibly accordingly.

AI implementation is complex and unpredictable. It involves multiple stakeholders and often coincides with other organizational changes, making it hard to measure the impact of AI alone. Researchers need to be flexible and prepared to adapt their research designs frequently, as early decisions may change.

Planning field experiments requires sufficient control over the treatment, in our case the introduction of AI technology, as well as over the research design. Otherwise, it is difficult to assign randomised experimental and control groups and to decide on the timing of measurements before and after the intervention (in our case: AI implementation). However, evaluating the introduction of AI tools and their impact in multinational companies can be very complex for several reasons, even for experts within the organisation:

1. Multinational companies typically have *complex management and organisational structures*. Hence, decisions often relate to a multitude of stakeholders both outside and inside the organisation, e.g., national politics or works councils.
2. The introduction of AI is often accompanied by *other changes in the organisational structure,* which can lead to disruptive factors in the scientific experiment. This makes it difficult for researchers to distinguish the impact of AI from the impact of (other) related organisational changes. In some cases, the associated other organisational changes take place before the AI projects are initiated - in anticipation of the start of the AI projects. Since in these cases the consequences occur before the cause has occurred, it is not always possible to clearly separate the effects and model relationships.

Due to this uncertain and complex process, researchers need to be aware that *the initial research design will likely need to be modified several times* during the research process. This is why we worked with the corporate partners at an early stage of their internal processes to develop our research design and measurements. However, this also meant that we had to start collaborating before the company-internal decision-making process was finalised. But as mentioned earlier, changes and delays in AI implementation processes are likely, and this implies that researchers need to be flexible in adapting their research strategy at short notice.

> **Lesson 2:** There are different phases of implementation that are not always comparable: In-depth knowledge of the AI implementation process is required

AI is introduced in phases, starting with pilot projects. However, differences across regions, self-selection in pilots, and AI's learning process make it hard to compare results. Researchers need deep knowledge of each phase and must work closely with companies to create effective research strategies.

The implementation of AI takes place in *various phases*. It often starts with a pilot project in one location, which is later expanded to other locations. In our original research plan, we use this staggered introduction of AI technology to identify experimental and control groups. However, the comparison between these groups can be affected for various reasons, making it *difficult to perform clean before/after measurements*.

1. The intervention can be modified in the initial phase, as *pilot projects* can take place in companies or teams that *voluntarily agree* to test the new technology (self-selection).
2. In addition, it can be difficult to use a staggered research design, as even within the same company, processes and available technology may be different *in different countries or regions*.
3. The various phases of technology introduction can also bring new challenges and hurdles. This is known as the *"last mile phenomenon"*[1]: Introducing a new technology can prove more difficult than trialling it in a more controlled environment during pilot projects and con-

1 This concept originally comes from logistics and indicates that the last mile of delivery to end-consumers is often most costly. However, the phenomenon can also be applied to innovation processes. See also two ai:conomics policy briefs on co-creation (concept and success story) at www.aiconomics.eu.

sequently requires different implementation strategies, making comparability between the different phases more difficult.

4. As *AI algorithms are constantly learning, the* performance of the AI can vary at different stages of deployment, making it difficult to accurately determine and differentiate between the impact of the initial "shock" that occurs during the initial deployment and subsequent "learning effects".

This means that the researchers must have *comprehensive knowledge* of the individual phases of the AI implementation process - it is not for nothing that the methodology used here is also referred to as "insider econometrics": Econometric models based on insider information, i.e., internal data and knowledge. This can only be achieved in a *co-creation process* in which the researchers work closely with the companies to develop the best possible research design and identification strategy (see also Cluster 2).

> **Lesson 3:** It is in the legitimate interest of companies to prevent the introduction of AI from being perceived as a shock for their employees: The use of fine-grained research tools is necessary

Companies try to prevent AI from being seen as a disruptive shock to employees by using gradual implementation and communication strategies. This makes it harder for researchers to measure clear before-and-after effects, so fine-grained research tools are needed to capture smaller, incremental changes.

In an ideal world, a crucial component of identifying causal effects requires researchers to analyse a sudden event or "shock" as it is usually called (that occurs either naturally or intentionally in the course of an intervention). This "shock" randomly affects otherwise comparable individuals or units. In addition, the larger the shock, the stronger the effects to be expected in the statistical model. When interventions occur as a shock, the effects are easier to model because the before and after can be easily distinguished and the effects of the shock can be measured clearly and promptly.

But this is the research world, and the corporate world is quite different. It is in the good interest of employees that companies will do: everything to ensure that employees do not experience the introduction of AI as a shock. The communication surrounding the introduction of the technology and the associated changes is planned, prepared, and accompanied in detail - the resulting effects cannot be identified as clearly as in the case of a shock. Companies favour situations in which the consequences of the introduction of technology can be predicted and controlled and try to avoid shocks as far as possible. Companies carefully plan the introduction of new technologies at the company's various locations and accompany them with change management campaigns, special communication concepts, etc. To mitigate the shock, the introduction often takes place in several and relatively small steps, which affects the number of companies and people involved - and makes it more difficult to clearly distinguish the "before" from the "after" and to model the effects.

For researchers, this means that they need to develop *fine-grained measuring instruments and methods that can* also be used to record smaller changes.

To summarise:
The process of AI implementation, as it is often found in corporate practice, makes it difficult to assign clear experiment and control groups and to identify effects. This includes:

- complex organisational structure and numerous interest groups;
- other organisational changes associated with the introduction of AI;
- different phases of the state of the art and pilot projects, which make it difficult to clearly define the individual phases of introduction and change:
- heterogeneous framework conditions in different locations;
- no clear "before" vs. "after";
- it is in the company's interest to avoid a shock.

Consequently, the researchers must:

- be flexible in adapting the research strategy if necessary;
- initiate a joint design process with the companies in order to develop the necessary comprehensive knowledge of the processes in the individual phases;
- develop fine-grained measures to assess minor changes;
- be in continuous contact with the companies to ensure a flexible adaptation to changing boundary conditions.

Cluster 2: Complexity of the co-creation process

Lesson 4: Trust is the basis for a successful co-creation process

Trust is the foundation of successful AI research collaboration, enabling access to insider data. It requires active investment in cooperation and communication, as well as flexibility from all parties involved.

We started our project with the preparation and elaboration of the research design and letters of intent (LOIs) for the funding body and companies. However, as already mentioned, AI implementation involves major changes and researching the consequences is only possible through access to insider data and information. This can only succeed if there is trust between all those involved in the project. In a way, trust is the basis, the ground on which the project can grow and become a success. Such trust does not develop automatically, but requires active investment in cooperation, shared values, and rules (see lesson 5), the recognition of sometimes divergent interests (see lesson 6) and good communication (see lesson 7), which must be maintained and cultivated throughout the entire duration of the project. Last but not least, co-creation requires flexibility and adaptability on all sides - including on the part of the funding body and the administering institutions.

Lesson 5: Co-creation requires continuous efforts in cooperation with the company

Co-creation in AI research demands close cooperation with companies, tailored research tools, and strong internal support. Engaging senior management is key to ensuring a successful collaboration.

To gain solid scientific insights into the real impact of AI on work and workers, *we need to understand exactly* what is going on in each of our cases. This requires working very closely with each company and co-designing the research (and the experiment) to tailor the research tools to each situation. This is necessary for three reasons: to get better information, to get better support, and to ensure that the project gets to be a win-win-win for all parties involved.

- In terms of *gathering better information*, we have found that we need detailed knowledge of the framework and technical aspects of the AI implementation process and its expected impact. This often requires knowledge of the technology and, e.g., the production process as such. In addition, *different experts in the organisation offer different perspectives*, all of which are necessary for a better understanding. This is not always easy, as the direct contacts in the company may also not know from whom they can obtain the necessary information.
- *Good support within the organisation is* needed to better organise the research process. For example, for our surveys we had to liaise directly with the employees in the different countries because we had to ensure the coherence of the research process in each country, e.g., in explaining the instructions for conducting the survey and ensuring links between the different waves. This requires a good support system within the organisation. At best, there are diagonal communication channels: from senior to middle management, to works councils, leaders and to employees.

To address these issues, *a highly engaged liaison pen at the executive level of* the company as well as *access to senior management* is mandatory to get things done in research-industry collaborations.

Lesson 6: Co-Creation requires recognising and balancing different interests

Co-creation requires balancing the differing interests of researchers, companies, and employee representatives. Open communication is key to addressing these differences and fostering trust.

In the co-creation processes, it is important to consider the different interests of stakeholders. While researchers primarily want to provide solid scientific evidence for the impact of AI and publish it in peer-reviewed journals, companies and corporate stakeholders have different (and sometimes conflicting) goals. Corporations may be more interested in demonstrating the impact of AI on productivity and using the results to make strategic decisions. Their works councils may be more interested in how AI affects employment opportunities or the well-being of employees and use the results to improve working conditions.

It is important to address these differences of interest openly and to create an atmosphere of trust in which these differences of interest can be discussed and balanced. Good communication is the key to this.

Lesson 7: Good communication can solve many problems

Effective communication bridges gaps between researchers and diverse stakeholders. Tailoring language and managing expectations are key to ensuring successful collaboration and aligning goals.

Good communication is often hindered by a *lack of common language* between researchers and the partners in the companies. There are many different stakeholders, such as management, internal project coordinators, engineers, data scientists, HR professionals, individual employees or works councils, who all have different perspectives and speak different "languages". Nevertheless, it is important to understand and communicate with all these stakeholders to conduct the research and *ensure overall coherence* in the information flow.

It is also important to *tailor the information to* the different levels of knowledge and information status. We have found that it is important to translate scientific language into terms that people can understand accurately and that additional efforts need to be made to optimise communication without compromising the underlying scientific needs.

Finally, we have realised that it is important to *deal* with *different expectations*. Companies want results that they can use for their own strategic decisions, but our research does not always deliver this, or not in a timely manner. This requires a continuous process of expectation management – and good communication.

To summarise:
To enable research in this area, well thought-out and intensive co-creation processes are required. This requires:

- creation of a fundamental relationship of trust between all parties involved;
- continuous cooperation in order to obtain sufficient information and support;
- a very dedicated liaison pin at the executive level of the organisation as well as access to higher management;
- 'diagonal' communication structures that can dock onto and be heard at all levels from management to employees and in all relevant bodies and committees;
- recognising and balancing differences of interest through openness and creating an atmosphere of trust;
- good communication to avoid misinterpretations, inconsistencies in the flow of information or false expectations.

Cluster 3: Methodological complexity

Lesson 8: It is often difficult to apply standard econometric tools that fit well with the complex processes of technology implementation such as AI, but a combination of tools and/or data can help

AI implementation is too complex for standard econometric tools alone. A combination of quantitative and qualitative methods is needed to fully understand its impact. Developing holistic approaches is key to future success in AI research.

The introduction of AI in companies is a complex process that *complicates the application of standard econometric tools*. Randomised control trials, double-blind trials, differences-in-differences and other quasi-experimental designs require identifiable processes, measurable effects and shocks that can be modelled (see Cluster 1). This is often not the case when implementing technologies such as AI. A diff-in-diff approach is still possible, but there are limitations as it is designed to measure the impact of a shock that the company primarily wants to minimise.

As several processes occur simultaneously, it is also challenging to isolate the impact of the introduction of AI from the other organisational changes associated with it. To gain a sound understanding of the technical details of AI, work organisation and the jobs affected, a *more holistic approach is* required, *combining state-of-the-art quantitative econometric tools with more qualitative approaches*. For example, in one of our cases, we found that the impact of AI implementation did not show up in the employee survey but did in the accompanying qualitative interviews and productivity study. Only by combining these different data sources and methods could we gain a better understanding of the processes.

Although *further approaches* still need to *be developed that successfully combine these different data sources and methods*, we firmly believe that this is the way forward.[2]

Lesson 9: We need to invest in developing metrics that capture the complexity and granularity of task changes and outcomes

Current metrics don't fully capture AI-driven task changes. New metrics are needed to reflect shifts in skills and task complexity. Working with company experts helps identify relevant KPIs, even if indirect, to measure AI's true impact.

Some of our field research has been hampered by the *lack of a good metric to study the impact on skill requirements and task changes*. Work tasks may not change on paper, but the context could change significantly, creating a different type of work that requires different skills. For example, there may be a shift *within the* task itself, with AI automating the simpler part of the task and leaving the more complex part to humans. With existing metrics, it is difficult to capture the complexity and granularity of these types of task changes.

One solution for developing such metrics would be to *analyse in depth* the tasks of the specific jobs of workers affected by AI and then measure, for example, the ratio of new to old tasks and the ratio of complex to routine tasks. Another solution to this problem could be to use *online job vacancies (OJV) to analyse the changes in skill requirements* before and after the introduction of AI technology, or to examine how *workers' profiles (in terms of education, skills, etc.)* change after the introduction of AI (here, however, longer-term changes are often to be expected). Analysing OJV data sometimes requires the use of machine learning techniques.

A second challenge is that the available KPI data is not always directly related to the outcome or level we are interested in. For example, AI can have *different effects on productivity and product quality.* Nevertheless, companies usually have more and better KPIs for productivity than for product quality. Similarly, it has sometimes proven difficult *to define KPI data* that directly links the changes caused by AI to outcomes at a specific level, e.g., when measures of consumer perception of product quality are only available at the team level rather than at the individual worker level or access to fine-grained data is not possible (due to technical and/or legal hurdles).

Working with experts in the company can help to find relevant KPIs for the results or level we are interested *in*. At first glance, certain data may not directly relate to the outcome or level we are interested in, but indirectly they do. This makes them useful from a scientific point of view. The search for such indirect indicators requires good and trusting cooperation with company insiders and technical experts.

Lesson 10: AI often only affects a small number of employees, but also look beyond the group of directly affected workers

AI often affects only a small group of employees, requiring alternative methods like vignette studies for small samples. It's important to also analyse spill-over effects beyond the directly impacted workers and use holistic data to understand AI's broader impact

In large companies often only a small number of employees are affected by the introduction of a *specific* AI, as a particular AI technology is usually targeted at a very *specific* part of the company's production process. As a result, the *number of employees directly affected for a particular use-case can be lower* than expected. For example, in some of our cases, we originally wanted to conduct surveys (before and after AI implementation), but due to the low number of observations after AI implementation, we decided to switch

2 A good example is the study by Hirvonen et al. (2023), which uses textual data and qualitative interviews to assess the impact of technology subsidies on employment and skill demand in Finnish manufacturing companies. See J. Hirvonen, A. Stenhammar and J. Tuhkuri (2023), New Evidence on the Effect of Technology on Employment and Skill Demand.

to vignette studies with stated preferences. This low number of observations also makes it challenging to study the impact of AI on workers, although methods such as synthetic control methods can help to increase the robustness of the results.

However, it is important to take a more holistic view and *look beyond the group of employees directly affected*. The introduction of AI will also have an impact on other parts of the organisation or lead to changes in the economy as a whole. These *spill-over effects* are equally important and should also be part of the analyses. As part of ai:conomics, the insider econometrics studies in and with companies are therefore supplemented by the analysis of register data (including social security data), which is available for representative groups of employees and provides insights into developments on the labour market in relation to AI.

To summarise:
Assessing the impact of AI implementation can be challenging for several reasons:

- standard econometric methods may not apply as they are based on identifiable processes and shocks to be modelled;
- lack of quantitative tools for a more holistic assessment of complex processes;
- lack of metrics to capture the complexity and granularity of changes in tasks and skills;
- lack of KPIs that can directly link the implementation of AI to a specific outcome or level;
- small number of employees affected.

This requires more investment in:

- holistic approaches that successfully combine different data sources and methods, in particular approaches that combine state-of-the-art quantitative econometric tools with more qualitative approaches;
- in-depth analyses of the tasks of the specific jobs of the employees affected by AI;
- use of machine learning tools to analyse changes in online job offers;
- working with experts in the organisation to develop relevant KPIs for the outcomes or levels we are interested in;
- analysing spill-over effects in other areas of the company or the economy.

Cluster 4 Legal and ethical complexity

Lesson 11: Data is often highly sensitive and requires strict protection regulations

AI research involves sensitive data that requires strict protection. Ethical reviews and legal agreements are essential to ensure compliance, and researchers must carefully balance data transparency with protecting companies' interests.

One of the main challenges for insider econometrics is that all data on the impact of AI on company results is extremely sensitive and data that is collected or used about individual employees, e.g., as part of field experiments (also referred to as 'human subject research') is anonymised, specially protected and may only be used to a (very) limited extent. Before such studies are carried out (before data collection even begins), an ethics council must therefore review the research project, not only for scientific journals that now require such a review as standard, but also for the stakeholders. The review requires that the data collection, analysis etc. is described in detail. This sometimes proves difficult in agreements with companies and relevant stakeholder groups, especially because processes, as described, are often dynamic and highly complex and information must be provided and decisions explained for the ethics board's review, which must first be developed together with the companies.

Besides, KPI data is often considered market-sensitive information that must not leave the organisation. And publications about the impact of AI implementation are also sensitive, either for market reasons (companies do not want to share certain results with competitors) or for internal political reasons (works councils may have different interests than management). This means that careful agreements need to be made on how to meet the legitimate interests of the company without losing sight of the goal of analysing the data in a robust way and publishing it in scientific journals and partly also public media (including social media). In addition, contracts must be negotiated and concluded (cooperation agreements, non-disclosure agreements and, for example, contracts on data transfer and utilisation).

These contracts also involve 'general' data protection issues within the meaning of the General Data Protection Regulation (GDPR), as we collect data on individual employees. Not only the European GDPR regulation must be complied with, but also the national and, if applicable, company-specific regulations. These different legal regulations must be handled with the utmost sensitivity and require legal experts on the side of the companies and as well as the research institutions who are familiar with the various framework conditions. Researchers must also be flexible and willing to familiarise themselves with the legal situation. Within the ai:conomics project, and our collaboration with the project and industry partners, we made sure to meet all required obligations.

To summarise:
The complexity of the legal framework for experimental research in and with corporate groups results from:

- the use of sensitive KPI data that requires a high level of data security (even if it not considered personal related data);
- a review by the responsible ethics council is necessary before data collection, for which the details of the research and data processing process must be written down, while agreements and decision-making processes between researchers and companies are often highly complex and dynamic and only develop over time;
- dealing with data protection regulations, which vary from country to country and from company to company.

This requires comprehensive legal and ethical expertise to deal with the various complexities.

Acknowledgments

We thank the experts who participated in the ai:conomics seminar „Lessons Learned“ on 21 June 2024, in particular Lajla Fetic (Applied AI Institute for Europe), Bart Golsteyn (Maastricht University), Paul Iske (Maastricht University), Judith Peterka (Federal Chancellery Germany) and Simon Wiederhold (IWH-Halle). We would also like to thank Michael Maier (Federal Ministry of Labour and Social Affairs), Simon Janssen and Michael Stops (IAB), Carolin Silbernagl (zukunft zwei GmbH) and Sander Dijksman, Danique Eijkenboom, Lara Fleck, and Andries de Grip (all ROA/University of Maastricht) for their comments and suggestions.