Technology, firms, and workers

Anna Gumpert

U Tübingen, CEPR, CESifo & IAW

Machine Learning in Science Conference

13 July 2026

European Research Council
Established by the European Commission

Funded by the European Union (ERC, ORGANDICT, 101076790).
Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Technology, firms and workers

New technologies have potential to **reshape employment**

- Widespread concern about unemployment due to automation
- Shifting labor demand creates winners and losers

Impact of technologies depends on **firms' decision to adopt and use technology**

My group's research:

- How does technology adoption reshape firms?
- How does firms' technology adoption affect workers?

Challenge I: Theory predicts countervailing effects

At least three countervailing effects of technology on employment

1. **Displacement effect**: technology replaces workers via automation

   $\Rightarrow$ Labor demand **declines**

2. **Productivity effect**: production costs decrease, so output increases

   $\Rightarrow$ Labor demand in non-automated tasks **increases**

3. **New tasks**: technology leads to new uses of labor

   $\Rightarrow$ Labor demand **increases**

Different effects affect **different workers**

Challenge II: Causal effects hard to estimate from observational data

It's hard (close to impossible) to run experiments with firms.

Regressing employee outcomes on firms' technology use does *not* yield causal effect.

Examples:

- New manager introduces new technology and outsources input production:
  Naïve regression attributes impact of outsourcing to technology use

- Firm does not find skilled employees and uses technology instead:
  Naïve regression attributes lower employment to technology use

This talk

1. Studying causal effect of information technology (IT) on firms and workers

2. Using ML to address structural breaks in administrative data

3. Using ML to advance research on broader economic questions

Studying causal effect of IT on firms and workers

Brewing, Gumpert, Steinwender, Antoni (2026): Unpacking the Black Box of ICT – Firm Organization and Worker Outcomes

Overview

Impact of **Enterprise Resource Planning (ERP) systems** on firm organization and worker outcomes

- ERP systems (e.g., SAP) are established information technology, but advances in AI promise to increase their efficiency
- Use **detailed data** and **theoretical model** to disentangle different possible effects on workers
- Use **quasi-experimental variation** in broadband access to address endogeneity of ERP adoption

Data

Establishment-level ICT data (Ci Technology Data Set, CiTDS), 2006-2019

- From private data provider, collected for ICT vendors
- Data on software (including ERP, ERP provider)

Linked employer-employee data (IAB), 1975-2019

- Universe of private-sector employees in Germany since 1975 from social security records
- Employment spells, daily wages, full-time/ part-time status, age, education, nationality, gender
- Detailed **occupation variable** with information on task complexity and leadership responsibility

German occupation classification system

Uniquely detailed German occupation classification

- Occupations classified with 5 digits
- 4th digit: 9 if manager or supervisor
- 5th digit: 4 types of task complexity and autonomy of task
  - 1: simple, not very complex, routine
  - 4: highly complex, requires highly specialized knowledge; **managers** always assigned level 4

⇒ Allows tracing changes to task complexity and autonomy *at employee level*

Example: accountants (class 722)

- 72212: assistant controller (e.g., Kaufmännischer Assistent)
- 72213: certified accountant (e.g., Bilanzbuchhalter/in)
- 72214: financial auditor (e.g., Rechnungsprüfer/in)
- 72294: head of controlling (e.g., Leiter/in - Controlling)

Empirical strategy: Combination of two approaches

Event-study analysis of ERP adoption (staggered difference-in-differences framework)

- Advantage: **precise** measure of technology adoption; **absence of pre-trends**
- Limitation: ERP adoption is firm decision, so despite absence of pre-trends effects not causal

$$y_{it} = \sum_{\substack{k=-K \\ k\neq -1}}^{L} \beta_k \cdot \mathbf{1}\{t - \text{ERP}_i = k\} + \alpha_i + \lambda_t + \varepsilon_{it}, \tag{1}$$

with y_{it} outcome of firm i in year t, α_i, λ_t: firm, year fixed effects

- Estimate group-time average treatment effects following Callaway & Sant'Anna (2021)
- Control group: not-yet-treated and never-treated firms (robust to not-yet-treated only)

Empirical strategy: Combination of two approaches

Intent-to-treat long-difference analysis using exogenous variation in cloud access

- Idea: most modern ERP systems cloud-based, however, cloud-based ERP requires stable internet connection with **high bandwidth for upload and download**

  "The natural barriers to full adoption include [...] speed/latency issues and reliance on telecommunication system providers." (KPMG 2012)

→ Exploit **exogenous variation** in cloud access for causal identification

- Source of variation: introduction of new telecommunications standard DOCSIS 3.0 that affected only regions with TV cables rolled out in the 1980s

In presentation: only event-study graphs

ERP adoption reduces employment

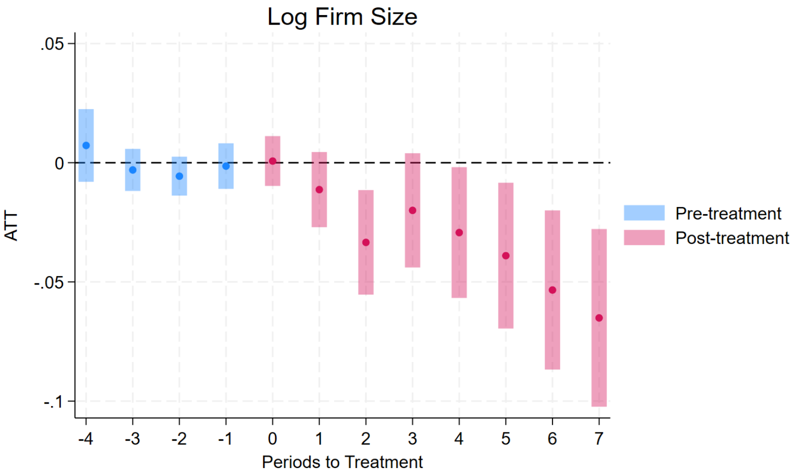
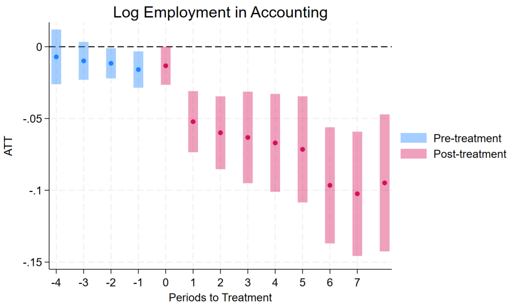
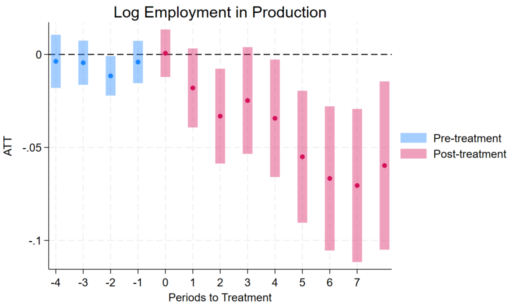


Figure: Event study: ERP adoption and log employment

Negative employment effects for support and production



(a) Support: Accounting



(b) Production

Figure: Event study: ERP adoption and log employment by function

reduced form

ERP adoption increases average task complexity

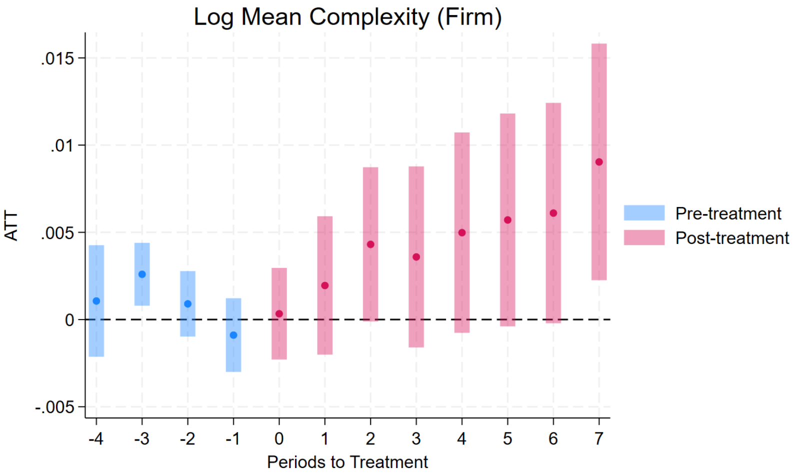
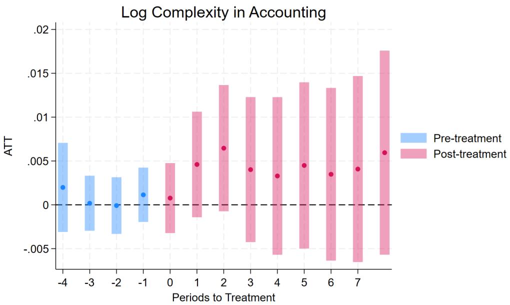
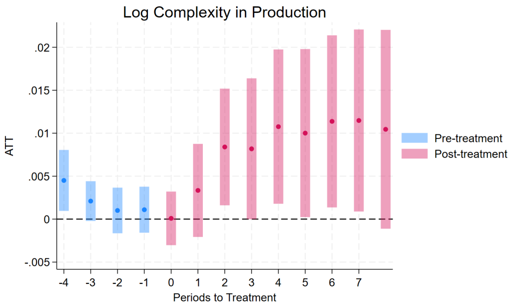


Figure: Event study: ERP adoption and log average task complexity

Complexity increases for production rather than support



(a) Support: Accounting



(b) Production

Figure: Event study: ERP adoption and task complexity by function

▸ reduced form

Mathematical model (sketch)

Knowledge hierarchy model (Garicano 2000; Caliendo & Rossi-Hansberg 2012; Gumpert et al. 2022)

- Production = problem solving process based on **labor and knowledge**
- Firm = CEO, production workers, possibly middle managers, support workers
  - Production workers input labor and some knowledge
  - CEO and middle managers use **knowledge** to help workers
  - Support workers reduce **information costs** (cost of knowledge) within firm

Assumption: IT substitutes for support workers and makes them more productive

- IT decreases # support workers
- IT renders **knowledge cheaper** relative to labor

$\Rightarrow$ IT decreases # production workers

IT increases worker knowledge (i.e., complexity)

Using ML to address structural breaks in administrative data

The issue: Structural break in occupation classification

New tasks: technology (and other factors) leads to new uses of labor

- Some jobs disappear, new jobs appear
- Statistical office **updates classification** (KldB 1988 $\rightarrow$ KldB 2010); provides mapping including information on "focal" transition

$\Rightarrow$ **Structural break** in administrative data: $m : m$ mapping from old, 3-digit classification to new, 5-digit classification

- Using observed new occupation for earlier years not appropriate: individual job-to-job transitions

Affects broad labor market questions, including e.g., estimation of gender wage gap

The idea: Use ML to improve mapping

Focus (for this talk): management/supervisory status (4th digit)

→ Goal: consistent individual-level measure of management/supervisory status, 2000–2020

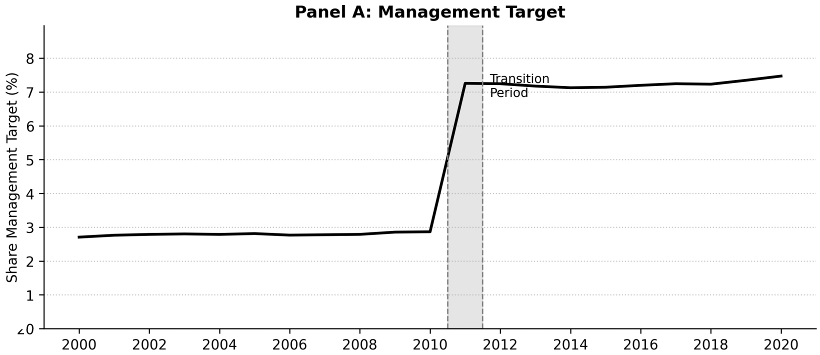


Figure: Share of employees with supervisory status, random 20% sample of German manufacturing workers

The idea: Use ML to improve mapping

Step 1: Prediction model (LightGBM)

- **Idea:** exploit 45 features observable on individual *beyond* their occupation classification: occupation code, wage relative to firm median, firm size/age interactions, ...
- Treat post-2012 codes as ground truth, combine two sources of information for pre-2012 status
- Trained on 2012–2018 ($\sim$7m obs.), tested on 2019–2020, person-disjoint split, lagged firm characteristics
- *Assumption:* relationship between characteristics and supervisory status stable across the break
- High-confidence predictions ($\textit{score} \geq 0.7$): precision 0.86, recall 0.85

The idea: Use ML to improve mapping

Step 2: Career consistency rules for remaining cases

- Manager in $t+1 \Rightarrow$ likely manager in t: firms rarely update reported codes for incumbent workers (persistence 87% in post-2012 data)
- Auxiliary: legacy classification identifies managers reliably, just too few of them
- *Assumption:* stable prevalence of 7.1–7.3%, consistent with flat legacy series and post-2012 level

Model provides high-precision core and ranking, rules complete the panel given assumptions

Result

Figure: Predicted and inferred supervisory status, 2000–2020

Using ML to advance research on broader economic questions

Who is an AI worker? Evidence from Matched Vacancy-Hires Hellsten (2026)

Fact 1

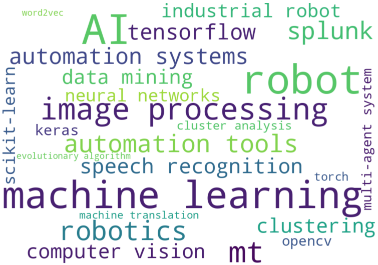
Workers with the right skills are key to development and uptake of AI.

Fact 2

There is a shortage of these workers.

...**But we don't know who they are?!**

Top AI skills (2023)



Policy: In order to train new AI talent, we need to know what traits define them

- Education?
- Experience?
- Something else?

Curious?!

⇒ Come to the **poster session**!

Intangible Investment in the Euro Area in Times of Uncertainty Saiz & Scrofani (in progress)

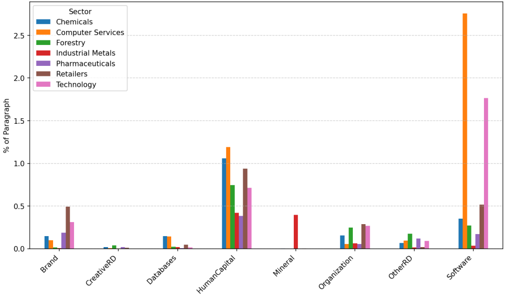
- Corporate investment increasingly **intangible**: software, R&D, branding, data, ...
- Intangible and tangible capital **structurally different**

⇒ How do different investment types respond to uncertainty and monetary policy?

The problem

We can barely measure intangible investments.

The solution: Use LLM to **extract information** on intangibles from corporate disclosures



Paragraphs classified in different intangible assets categories

Curious?! ⇒ Come to the **poster session**!

Bulky Offshoring

Eppinger, Gumpert, Heiland, Hellsten, Koch (in progress)

Technology adoption and firms' offshoring decisions are **closely intertwined**.

- Technology **enables** offshoring: modern IT facilitates management at a distance
- Technology and offshoring are **competing options** to cut labor costs

Bulky Offshoring

Eppinger, Gumpert, Heiland, Hellsten, Koch (in progress)

Administrative data provide information on

- Employee outcomes: employment status, wages; occupation
- Imports at product level

Theory predicts **countervailing effects** of offshoring on employees:

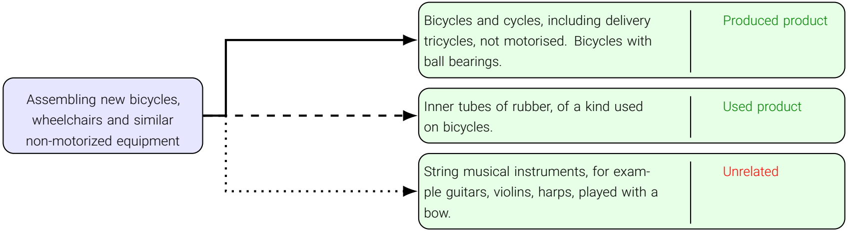
- Firm offshores product *produced* by worker: **negative** employment effect
- Firm offshores product *used* by worker as input: **positive** employment effect

Bulky Offshoring

Eppinger, Gumpert, Heiland, Hellsten, Koch (in progress)

Idea:

Use NLP, job and product descriptions to differentiate offshoring of *produced and used* products



Curious? ⇒ Let's talk over coffee!

Thanks!

How to find us:

- Anna Gumpert: anna.gumpert@uni-tuebingen.de, https://annagumpert.wordpress.com/

- Mark Hellsten: mark.hellsten@uni-tuebingen.de, https://sites.google.com/view/markhellsten/home

- Stefania Scrofani: stefania.scrofani02@gmail.com, https://stefaniascrofani.github.io/