STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, April 2026, pp 3059–3084.
Published online in International Academic Press (www.IAPress.org)

Causal Inference in Econometrics Using Machine Learning: Estimating the Effect of AI and Automation Adoption on Firm Productivity in Europe

Md. Shahidul Islam [1,*], Md. Ashikquer Rahman [2], Akbar Ali Hossain [3]

[1]*Faculty of Sciences & Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh*
[2]*Faculty of Business & Economics, Bangladesh University of Business and Technology, Dhaka, Bangladesh*
[3]*School of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia*

Abstract Artificial intelligence and automation are becoming central to how European firms work, compete, and organize their production. But despite the rapid growth of these technologies, there is still a key question: does adopting AI genuinely make firms more productive, or are already-productive firms simply more inclined to adopt it? This study addresses that question by combining established econometric methods with newer causal machine learning techniques. Using a large panel of European firms from 2010 to 2023, built from Orbis and EU KLEMS, AI adoption is identified through both investment measures and text based disclosure indicators. Across multiple empirical approaches including fixed effects, difference in differences, and instrumental variable models the results consistently show productivity gains of roughly 3% to 6% among AI adopting firms. Double Machine Learning produces a similarly robust estimate of around 4.5%. Event study evidence further indicates no pre adoption improvements, with productivity gains emerging gradually afterward. The effects, however, are uneven. Larger firms, those with more advanced digital systems, and firms employing a higher share of skilled workers benefit noticeably more from AI adoption. In contrast, firms lacking strong digital foundations or sufficient human capital see smaller gains. The results indicate that adopting AI does boost firm productivity, but the size of the benefit depends heavily on the presence of complementary skills and digital infrastructure.

Keywords Causal Inference, Machine Learning for Causality, Econometric Modeling, AI and Automation Adoption, Firm-Level Productivity Analysis.

AMS 2010 subject classifications 62M10, 68T10

DOI: 10.19139/soic-2310-5070-3212

1. Introduction

Across Europe and beyond, the expansion of artificial intelligence (AI) and automation technologies is often described as the rise of a new general-purpose technology (GPT) capable of redefining production and firm competitiveness [1, 2]. European firms have rapidly adopted AI-driven systems for predictive analytics, supply chain optimization, and process automation, supported by initiatives such as the EU Digital Strategy and national-level AI policies. However, it remains an open empirical question whether AI adoption causally improves firm productivity, or whether more productive firms simply self-select into adopting these technologies. This empirical challenge arises for two main reasons. First, technology adoption is endogenous firms with higher productivity, better management, or greater access to capital are more likely to adopt AI [3, 4]. This leads to selection bias and confounds causal interpretation. Second, the impact of AI is likely heterogeneous across industries, firm size, and organizational capabilities [1, 5]. Traditional econometric models such as ordinary least squares (OLS), difference-in-differences (DID), and instrumental variables (IV) have been widely used to estimate causal effects [6], but they

*Correspondence to: Md. Shahidul Islam (Email: sshahid01921@gmail.com). Faculty of Sciences & Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

rely on strong linearity, additivity, and homogeneity assumptions. Recent methodological advances have introduced causal machine learning (CML) as a powerful alternative. CML methods such as Double Machine Learning (DML) [7], Causal Forests [8], and Meta-learners (e.g., T-learner, S-learner, X-learner) [9] integrate identification strategies from econometrics with flexible, data-driven function approximation from machine learning. These methods employ regularization, cross-fitting, and sample splitting to correct for overfitting and provide valid inference. Importantly, they enable estimation of not only the average treatment effect (ATE) but also conditional average treatment effects (CATE), capturing heterogeneity in technology impacts.

In this paper, we apply both traditional econometric techniques and state-of-the-art CML methods to estimate the causal effect of AI and automation adoption on firm productivity in Europe. We construct a comprehensive panel by combining firm-level data from Orbis with industry-level productivity and ICT capital data from EU KLEMS. We identify AI adoption using indicators such as AI-related intangible assets, automation capital expenditures, and text-based measures from annual reports [10, 11]. Beyond applying existing methods, this study introduces a unified empirical framework that integrates panel fixed-effects identification with modern causal machine learning estimators to jointly estimate average and heterogeneous treatment effects. The combination of econometric identification, cross-fitted machine learning nuisance estimation, and firm-level heterogeneity analysis represents a methodological advancement in applied causal inference for technology adoption research.

The objectives of this study are fourfold. First, it aims to estimate the causal effect of AI and automation adoption on firm productivity in Europe. Second, it seeks to compare traditional econometric methods with causal machine learning estimators in assessing these productivity effects. Third, the study examines heterogeneous treatment effects across industries and firm characteristics, including firm size and technological intensity. Finally, it aims to derive policy implications to support AI-driven digital transformation in Europe. To address these objectives, the study is guided by four research questions.

1. Does AI and automation adoption causally improve firm productivity?
2. Do causal machine learning methods yield different or more robust estimates than traditional econometric approaches?
3. How do productivity effects vary across industries, firm size, skill intensity, and digital maturity?
4. What policy lessons can be drawn for inclusive AI-driven growth in Europe?

The remainder of the paper is organized as follows. Section 2 reviews related literature on causal inference and technology adoption. Section 3 outlines the methodological framework, including econometric and machine learning estimators. Section 4 describes the data sources and variable construction. Section 5 presents the empirical results. Section 6 discusses policy implications and limitations. Finally, Section 7 concludes the paper.

2. Literature Review

2.1. Traditional Econometric Approaches to Causal Inference

Causal inference in econometrics has historically been grounded in the potential outcomes framework, also known as the Rubin Causal Model, which formalizes treatment effects as comparisons between observed and counterfactual outcomes [12]. Under this framework, the fundamental problem of causal inference arises because only one potential outcome is observed for each unit, requiring researchers to rely on assumptions and statistical methods to recover causal effects. Traditional techniques such as Ordinary Least Squares (OLS) have been widely used to estimate causal relationships; however, OLS relies on assumptions of exogeneity and linearity, which often do not hold in observational data where selection bias and omitted variable bias are present [13].

To address endogeneity, econometricians developed Instrumental Variables (IV) methods, which identify causal effects by exploiting external sources of variation that affect treatment but not the outcome directly [14]. IV has been influential in labor economics and public finance, but valid instruments are difficult to find, and weak instruments can severely bias estimates. Another widely used identification strategy is Difference-in-Differences (DID), which exploits temporal variation in treatment exposure across groups. Seminal studies such as Card and Krueger's minimum wage research exemplify DID's effectiveness, but the method assumes parallel trends, which may not hold in heterogeneous or dynamic environments [15]. In addition, Regression Discontinuity Designs (RDD) leverage cutoff-based assignment rules to approximate randomized experiments, but require strict continuity assumptions and sufficient density around the threshold [16].

Other approaches such as Propensity Score Matching (PSM) aim to balance observable covariates between treated and control groups by estimating the probability of treatment and matching similar firms [17]. While matching methods reduce selection bias on observables, they cannot adjust for unobserved confounders and perform poorly in high-dimensional settings. A common limitation across these traditional methods is the reliance on strong functional form assumptions and limited capacity to model heterogeneity. As modern datasets become larger and more complex, with nonlinear relationships and numerous covariates, traditional econometric tools often struggle to produce unbiased and robust causal estimates.

2.2. Machine Learning for Causal Inference

Machine learning (ML) methods have emerged as powerful tools for prediction due to their ability to model nonlinear relationships and interactions using flexible function approximation [18]. However, standard ML algorithms optimize predictive accuracy rather than causal identification, which limits their direct applicability to causal inference. A key breakthrough occurred when researchers recognized that ML could be incorporated into econometric frameworks to estimate nuisance components such as propensity scores or baseline outcome functions, while still preserving the asymptotic properties required for valid inference [19]. One of the most influential developments is Double Machine Learning (DML), introduced by Chernozhukov et al. [7].

DML combines orthogonalization (to remove bias from nuisance parameter estimation) with cross-fitting (to prevent overfitting), enabling consistent estimation of treatment effects in high-dimensional settings. DML has been widely applied in economics, labor markets, and policy evaluation, demonstrating improved robustness compared to traditional econometric approaches. Another major innovation is Causal Forests, developed by Wager and Athey [8], which extend Random Forests to estimate heterogeneous treatment effects (HTE). Causal Forests recursively partition the data to minimize treatment effect variance, capturing complex interactions between covariates and enabling personalized treatment effect estimates at the firm level. While Random Forests prioritize prediction accuracy, Causal Forests explicitly target causal heterogeneity. Additionally, Meta-learners such as T-learners, S-learners, and X-learners provide flexible frameworks for combining machine learning models with causal effect estimation [9].

Recent advances in machine learning demonstrate strong capability in detecting nonlinear patterns, structural shifts, and latent dynamics in complex time series systems [20, 21, 22], providing methodological foundations for modeling dynamic treatment effects. Building on these predictive and structural learning approaches, this study extends their principles to causal econometrics to estimate the impact of AI and automation adoption on firm productivity in Europe under high-dimensional and heterogeneous settings [23, 24].

These methods decompose causal estimation into prediction tasks, enabling the use of any supervised learning algorithm (e.g., boosted trees, neural networks, elastic nets). The X-learner is particularly effective in observational settings with imbalanced treatment groups. More recently, neural network–based causal models and Bayesian Additive Regression Trees (BART) have demonstrated strong performance in modeling nonlinearities and capturing heterogeneous effects [25]. Despite their advantages, causal ML methods require large samples, careful hyperparameter tuning, and rigorous validation to avoid overfitting. Interpretability may also be challenging with complex models. Nevertheless, the integration of econometric identification strategies with ML flexibility has created a powerful new paradigm in empirical research commonly referred to as causal machine learning (CML) [7, 19].

2.3. Technology Adoption, AI, and Firm Productivity

The relationship between technology adoption and productivity has been widely studied in the economics of innovation. Early work by Solow highlighted the “productivity paradox,” where advances in computing did not immediately translate into measurable productivity gains [26]. Subsequent studies by Brynjolfsson and Hitt showed that Information and Communication Technology (ICT) investments significantly increase firm productivity, especially when complemented by organizational restructuring and managerial practices [27]. ICT has been identified as a key driver of productivity growth across industries and countries. More recent research views artificial intelligence (AI) as a new general-purpose technology (GPT) with the potential for wide-ranging productivity impacts [28].

Studies suggest that AI enhances automation, predictive accuracy, and innovation capabilities, but its benefits vary across firms and sectors. Cockburn, Henderson, and Stern argue that AI serves as a "method of invention," raising research productivity and accelerating innovation [29]. Brynjolfsson, Rock, and Syverson propose a "productivity J-curve," predicting that short-term productivity may stagnate or decline during the adjustment phase before long-term gains materialize [28]. Firm-level empirical evidence on AI adoption is expanding but remains limited. Several studies using U.S. and international datasets find positive correlations between AI adoption and productivity, profitability, and innovation outcomes [27, 29].

However, most rely on correlational approaches, preventing causal interpretation. Furthermore, heterogeneous treatment effects, for example, across firm size, digital intensity, or workforce skills are rarely quantified. In the European context, databases such as Orbis provide rich firm-level financial and ownership data, while EU KLEMS offers industry-level productivity and ICT capital measures, making them well suited for studying AI adoption and productivity [30]. OECD and European Commission reports document substantial cross-country variation in digital infrastructure, technology adoption, and workforce skills across EU member states [31]. These differences likely influence both the propensity to adopt AI and the magnitude of its productivity effects. Despite growing interest, two major gaps remain in the literature. First, limited causal evidence exists on the productivity impact of AI adoption at the firm level, especially using European data. Second, existing studies rarely apply causal machine learning methods to uncover heterogeneous effects. Addressing these gaps, this paper employs both traditional econometric and CML methods to estimate the causal effect of AI adoption on firm productivity using a cross-country panel of European firms. This approach responds to recent calls for more rigorous causal identification and the use of flexible, data-driven tools in the study of technology and productivity [19, 28].

2.4. Literature Gaps and Contributions

- **Causal identification and endogeneity:** Existing studies on technology adoption and productivity commonly rely on conventional econometric approaches such as OLS, DID, and IV; however, causal identification remains challenging due to endogeneity, selection bias, and strong parametric assumptions [6, 13, 14, 15, 17]. This study strengthens causal inference on AI adoption and firm productivity by combining traditional econometric identification strategies with modern causal machine learning estimators [7, 19].
- **Methodological advancement (econometrics vs. causal ML):** Traditional econometric models provide interpretable causal frameworks but are limited in capturing nonlinearities and complex interactions, while standard machine learning methods are primarily designed for prediction rather than causal inference [18, 19]. This study contributes by systematically comparing classical econometric estimators with state-of-the-art causal machine learning approaches, including Double Machine Learning and Meta-learners [7, 8, 9].
- **Heterogeneous treatment effects:** Much of the existing empirical literature focuses on average productivity effects, although the impact of AI is expected to vary substantially across firms and industries [1, 5, 28]. This study addresses this gap by estimating heterogeneous treatment effects using Causal Forests and Meta-learners and by identifying key drivers of variation across industries, firm size, and technology intensity [8, 9].
- **European evidence and policy relevance:** While recent research highlights AI as a general-purpose technology with the potential to enhance productivity, firm-level causal evidence remains limited, particularly in the European context [28, 29, 31]. This study contributes new European evidence by combining Orbis firm-level data with EU KLEMS industry-level productivity and ICT capital measures, generating targeted implications for digital transformation and workforce upskilling policies [30, 31].

2.5. Testable Hypotheses

Building on the foregoing review of the literature and the identified gaps concerning causal identification, methodological limitations, and unobserved heterogeneity in the productivity effects of AI adoption, this study formulates a set of testable hypotheses.AI and automation adoption is expected to have a positive causal effect on firm productivity (H1), and this positive effect is hypothesized to remain statistically significant across both traditional econometric estimators and causal machine learning approaches (H2). Given the uneven diffusion of AI across sectors and firms, the productivity effects of AI adoption are further hypothesized to be heterogeneous across industries (H3) and firm size categories, with larger firms benefiting more than smaller firms (H4). Finally, consistent with complementarity theory, firms with higher human-capital intensity and greater digital maturity are expected to obtain larger productivity gains from AI and automation adoption (H5).

3. Methodology

3.1. Causal Framework

The causal effect of AI and automation adoption on firm productivity is analyzed under the potential outcomes framework, which defines treatment as the adoption of AI technologies by a firm. Let $D_i = 1$ if firm i adopts AI, and $D_i = 0$ otherwise. Each firm has two potential outcomes: $Y_i(1)$, the productivity if it adopts AI, and $Y_i(0)$, the productivity if it does not adopt AI [29]. The individual treatment effect is defined as:

$$\tau_i = Y_i(1) - Y_i(0).$$

However, only one of the two potential outcomes is observed for each firm. Therefore, causal inference requires assumptions or statistical methods to estimate average treatment effects. The primary estimands of interest are the Average Treatment Effect (ATE):

$$ATE = \mathbb{E}[Y_i(1) - Y_i(0)],$$

and the Conditional Average Treatment Effect (CATE):

$$CATE(x) = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x],$$

where X_i represents firm characteristics such as size, industry, capital intensity, or digital maturity. Estimating CATE allows us to capture heterogeneous treatment effects, which are likely important in the context of AI adoption.

A key challenge in estimating ATE or CATE is selection bias, which occurs when AI adoption is correlated with unobserved productivity determinants. This violates the unconfoundedness assumption:

$$(Y_i(0), Y_i(1)) \perp D_i \mid X_i,$$

which rarely holds in observational data, motivating the use of advanced econometric and machine learning methods to achieve valid identification.

3.2. Econometric Approaches

A standard approach to estimating the effect of AI adoption is Ordinary Least Squares (OLS) with control variables:

$$Y_{it} = \alpha + \beta D_{it} + \gamma' X_{it} + \varepsilon_{it},$$

where Y_{it} is the productivity of firm i at time t, and X_{it} includes observable controls. While simple, OLS relies on strong linearity and exogeneity assumptions, which may not hold.

To address unobserved heterogeneity, panel data models with fixed effects are commonly used:

$$Y_{it} = \alpha_i + \lambda_t + \beta D_{it} + \gamma' X_{it} + \varepsilon_{it},$$

where α_i captures time-invariant firm effects and λ_t captures time effects. This controls for unobserved firm-specific productivity differences but cannot account for time-varying unobserved confounders.

Instrumental Variables (IV) and Two-Stage Least Squares (2SLS) estimators address endogeneity by exploiting exogenous variation in AI adoption. For example, regional AI policy incentives or digital infrastructure rollout may serve as instruments. IV provides consistent estimates if instruments satisfy relevance and exclusion restrictions [14], but valid instruments are often difficult to obtain, and weak instruments can bias results.

Difference-in-Differences (DID) compares changes in productivity for firms that adopt AI (treated) and those that do not (control) before and after adoption:

$$Y_{it} = \alpha + \beta(D_i \times Post_t) + \gamma' X_{it} + \varepsilon_{it}.$$

DID assumes parallel productivity trends in the absence of treatment [15]. Extensions such as staggered DID and synthetic control methods can improve robustness, but DID remains limited in capturing nonlinear and heterogeneous effects across firms.

Overall, traditional econometric methods provide clear identification strategies but lack flexibility in modeling complex relationships and interactions prevalent in AI adoption settings.

3.3. Causal Machine Learning Approaches

Causal Machine Learning (CML) methods integrate econometric principles with machine learning flexibility to estimate treatment effects in high-dimensional and heterogeneous environments. These methods address both prediction and inference by using ML to estimate nuisance functions while preserving the orthogonality conditions needed for valid causal inference [19].

3.3.1. Double Machine Learning (DML) :
Double Machine Learning (DML), proposed by Chernozhukov et al. [7], estimates treatment effects using three key steps:

1. Estimate nuisance functions $\hat{m}(X) = \mathbb{E}[Y \mid X]$ and $\hat{p}(X) = \mathbb{E}[D \mid X]$ using ML models.
2. Compute residuals:
$$\tilde{Y} = Y - \hat{m}(X), \qquad \tilde{D} = D - \hat{p}(X).$$
3. Regress $\tilde{Y}$ on $\tilde{D}$ to obtain $\hat{\beta}$.

DML employs cross-fitting to prevent overfitting and orthogonalization to remove bias from nuisance parameter estimation. It is robust in high-dimensional settings and provides valid confidence intervals.

3.3.2. Causal Forests :
Causal Forests, introduced by Wager and Athey [8], extend Random Forests to directly estimate treatment effect heterogeneity. Instead of minimizing prediction error, Causal Forests split the data to maximize treatment effect variation. The estimated treatment effect for a firm with characteristics x is:

$$\hat{\tau}(x) = \mathbb{E}[Y \mid D = 1, X = x] - \mathbb{E}[Y \mid D = 0, X = x].$$

Causal Forests are nonparametric and capture complex interactions, making them well suited to uncover which types of firms benefit most from AI adoption.

3.3.3. Meta-learners :
Meta-learners provide flexible frameworks that transform causal estimation into prediction tasks using ML models [9]. The **T-learner** trains separate models for treated and control units:

$$\hat{\tau}_T(X) = \hat{\mu}_1(X) - \hat{\mu}_0(X).$$

The **S-learner** fits a single model with treatment included as a feature:

$$\hat{\tau}_S(X) = \hat{\mu}(X, 1) - \hat{\mu}(X, 0).$$

The **X-learner**, especially effective under treatment-group imbalance, proceeds in four steps:

1. Estimate $\hat{\mu}_1(X)$ on treated firms.
2. Estimate $\hat{\mu}_0(X)$ on control firms.
3. Compute imputed treatment effects:

$$\hat{\tau}_1 = Y_1 - \hat{\mu}_0(X_1), \qquad \hat{\tau}_0 = \hat{\mu}_1(X_0) - Y_0.$$

4. Combine estimates using propensity scores:

$$\hat{\tau}_X(X) = g(X)\hat{\tau}_0(X) + (1 - g(X))\hat{\tau}_1(X).$$

Meta-learners can incorporate neural networks, gradient boosting, or other ML algorithms, performing well in high-dimensional environments. However, they require careful tuning and transparency to ensure interpretability.

3.4. Identification Strategy

Our identification strategy integrates traditional econometric assumptions with the flexibility of machine learning to estimate the causal effect of AI adoption. We define the treatment variable $D_{it} = 1$ if firm i adopts AI in year t, and $D_{it} = 0$ otherwise. AI adoption is identified using indicators such as AI-related intangible assets, automation-related capital expenditures, and text-mined disclosures from annual reports or corporate filings. We construct treatment and control groups using propensity score estimation via ML models to balance observables. To address endogeneity from unobserved factors, we explore two strategies:

1. Panel Fixed Effects: controlling for time-invariant firm heterogeneity.
2. Instrumental Variables or Policy Shocks: using variation in regional AI incentives or digital infrastructure rollout as instruments.

We then estimate treatment effects using both traditional econometric models (FE, IV, DID) and CML methods (DML, Causal Forests, Meta-learners). This hybrid approach allows us to compare estimator performance and assess robustness across identification strategies.In the current implementation, instrumental variables were applied within the traditional IV/2SLS framework to address endogeneity concerns, while DML was used primarily as an orthogonalized selection-on-observables estimator. Importantly, the IV instruments were not directly embedded into the Causal Forest or baseline DML estimators (i.e., we did not implement a formal DML-IV estimator). Instead, the IV estimates serve as an external robustness benchmark, confirming that causal conclusions remain consistent under alternative identification assumptions

3.5. Machine Learning Implementation Details

To enhance reproducibility and ensure transparency in the causal machine learning framework, we provide additional implementation details regarding the machine learning models used to estimate the nuisance components in the Double Machine Learning (DML) and meta-learning estimators. In particular, all supervised learners employed for estimating the outcome function $m(X)$ and the treatment assignment function $p(X)$ including Random Forests, Gradient Boosting Machines (XGBoost), LASSO regression, and Neural Networks were systematically tuned using rigorous cross-validation procedures.

Specifically, we applied five-fold cross-validation throughout all nuisance estimation steps. This procedure ensured that hyperparameters were selected based on out-of-sample predictive performance rather than in-sample fit, thereby reducing the risk of overfitting. Hyperparameter tuning was conducted using grid-search strategies over key model parameters, including the number of trees, maximum depth, and minimum leaf size for Random Forests; learning rates, boosting iterations, and subsampling rates for Gradient Boosting models; and penalty parameters for LASSO regularization. For Neural Networks, tuning focused on the number of hidden layers, neuron size, dropout rates, and optimization learning schedules.

The nuisance functions $m(X)$ and $p(X)$ were therefore estimated flexibly in high-dimensional covariate space, allowing nonlinearities and interaction effects among firm characteristics such as size, industry classification, ICT intensity, human capital composition, and digital readiness to be captured without imposing restrictive functional form assumptions. This flexibility is particularly important in observational firm-level data, where treatment assignment (AI adoption) is likely driven by complex selection mechanisms.

To further ensure valid causal inference, we adopted the standard cross-fitting and sample-splitting strategy proposed by [7]. In each fold, nuisance models were trained on one subsample and evaluated on a held-out subsample, so that residualized outcomes and treatments were constructed using predictions generated out-of-sample. This orthogonalization step guarantees that estimation errors in nuisance functions do not bias the causal parameter of interest.

Formally, the orthogonalized residuals were computed as:

$$\tilde{Y} = Y - \hat{m}(X), \qquad \tilde{D} = D - \hat{p}(X),$$

and the causal treatment effect was estimated via the DML orthogonal moment condition:

$$\hat{\tau}_{DML} = \frac{\mathrm{Cov}(\tilde{Y}, \tilde{D})}{\mathrm{Var}(\tilde{D})}.$$

This procedure provides asymptotically unbiased and $\sqrt{n}$-consistent estimates even when nuisance components are learned using complex machine learning algorithms. Standard errors were computed using heteroskedasticity-robust variance estimators consistent with DML asymptotic theory. Finally, robustness was assessed by re-estimating the baseline ATE across alternative learners and tuning specifications.

3.6. Estimation and Evaluation

We estimate the Average Treatment Effect (ATE) using several econometric approaches, including OLS, Fixed Effects (FE), Instrumental Variables (IV), Difference-in-Differences (DID), and Double Machine Learning (DML). To capture treatment heterogeneity, we estimate Conditional Average Treatment Effects (CATE) using Causal Forests and Meta-learners. Model performance is evaluated based on point estimates and statistical significance, model fit metrics such as mean squared error (MSE), robustness checks across alternative specifications, and heterogeneity analyses across firm size, industry, and digital intensity. Additionally, placebo tests, pre-trend analyses, and sensitivity checks are conducted to validate identification assumptions and ensure reliability.

Comparing econometric estimators with causal machine learning approaches helps assess whether flexible, ML-based methods can reveal additional causal relationships and heterogeneous effects that traditional econometric models may miss. This comparison is scientifically valuable for several reasons. First, conventional econometric methods provide clear identification strategies and highly interpretable parameters, but they often depend on strong assumptions about functional form and treatment homogeneity. Second, causal machine learning methods relax many of these restrictions, accommodating nonlinearities and high-dimensional confounding while still enabling valid inference through tools such as orthogonalization and cross-fitting. Third, evaluating results across both frameworks allows researchers to test the robustness of findings across methodological paradigms, improving credibility and uncovering treatment effect heterogeneity that standard models are not designed to detect. Together, this combined approach strengthens causal conclusions and contributes to more rigorous applied econometric research.

4. Data and Variable Construction

4.1. Data Sources

This study combines firm-level financial data from Orbis (Bureau van Dijk) with industry-level productivity and technology data from EU KLEMS. Orbis provides detailed information on European firms, including financial statements, employment, assets, ownership structure, and industry classification using NACE Rev. 2 codes. It is widely used in empirical research on corporate performance and innovation due to its extensive coverage of both public and private firms across European countries. EU KLEMS offers industry-level measures of labor productivity, total factor productivity (TFP), capital input, ICT capital, and labor quality for European economies.

The dataset is designed to analyze productivity and technological change across industries and countries, making it ideal for studying the relationship between digital technology and productivity. We use the 2019 release, updated through 2023 where available. The sample period covers 2010–2023, a period characterized by rapid digital transformation, increased AI and automation adoption, and multiple EU policy initiatives supporting digitalization. The geographic scope includes major EU economies such as Germany, France, Italy, Spain, the Netherlands, Sweden, and selected Eastern European countries, ensuring substantial variation in technology adoption and institutional context. Orbis and EU KLEMS are highly complementary: Orbis offers micro-level firm heterogeneity, while EU KLEMS captures macro-level industry productivity and ICT intensity. By merging these datasets, we construct a unique panel that allows for rigorous causal inference on AI adoption and productivity at the firm level.

4.2. Data Integration and Panel Structure

The integration of Orbis and EU KLEMS data proceeds in several steps. First, each firm in Orbis is assigned to an industry using 4-digit NACE Rev. 2 classification. EU KLEMS data are aggregated or disaggregated where necessary to match the same industry codes. This allows us to attach industry-level productivity and ICT measures to each firm-year observation. We construct a balanced panel of firms observed for at least five consecutive years to ensure valid estimation of firm fixed effects and pre-trends. To minimize the influence of extreme outliers, financial variables are winsorized at the 1st and 99th percentiles. All monetary values are deflated using country-specific GDP deflators and converted to a common currency (EUR) using Purchasing Power Parity (PPP) conversion factors. We exclude firms from industries where AI adoption is structurally less relevant (e.g., agriculture, mining) and focus on manufacturing, ICT, finance, professional services, transportation, and retail, which are major adopters of automation and AI. Missing values are addressed using firm-level interpolation for short gaps and listwise deletion for critical variables. After cleaning, the final dataset includes approximately 50,000 firms across 10 countries and 14 years, yielding over 500,000 firm-year observations. This panel structure allows us to exploit within-firm variation in AI adoption over time, control for unobserved heterogeneity using fixed effects, and apply causal machine learning techniques that require large sample sizes.

4.3. Variable Construction

4.3.1. Dependent Variable: Firm Productivity :

We employ two primary measures of firm productivity. The first is labor productivity, defined as

$$LP_{it} = \frac{\text{Value Added}_{it}}{\text{Employees}_{it}},$$

where value added is calculated as revenue minus intermediate inputs. Labor productivity is consistently available across firms and therefore serves as the baseline indicator.

The second measure is total factor productivity (TFP), derived either from values reported in Orbis or estimated using a Cobb–Douglas production function that incorporates capital and labor inputs. TFP captures a firm's overall efficiency beyond the accumulation of production factors and is widely used in empirical productivity research. While labor productivity provides broad coverage, TFP is employed as a robustness measure to validate the consistency of our findings.

4.3.2. Treatment Variable: AI/Automation Adoption :

The treatment variable, AI and automation adoption, is constructed using several complementary proxies because AI use cannot be directly observed through a single metric. First, we measure AI-related intangible assets, including capitalized software, algorithmic systems, and AI-oriented patents, identifying a firm as an adopter when these intangibles exceed a specified proportion of total assets. Second, automation capital expenditures are captured using investments in robotics, machinery, and automated production technologies extracted from fixed-asset categories in Orbis. Third, we employ a text-based AI disclosure indicator by applying NLP techniques to annual reports and searching for terms such as "machine learning," "automation," and "algorithm," following established approaches in finance and innovation research. Finally, we integrate industry-level AI intensity from EU KLEMS and OECD

Table 1. Validation and Robustness of the AI Adoption Measure

Specification	Validation Approach	Coefficient	Std. Error	p-value	R^2
Baseline DML (Binary AI)	Composite AI adoption measure (baseline specification)	0.045	0.011	0.001	0.412
Threshold 3% (Binary AI)	Alternative adoption cutoff (lower threshold)	0.042	0.010	0.002	0.409
Threshold 7% (Binary AI)	Alternative adoption cutoff (higher threshold)	0.047	0.012	0.001	0.415
Threshold 10% (Binary AI)	Stringent adoption definition (high threshold)	0.044	0.013	0.003	0.410
Continuous AI Intensity	Disaggregated treatment: continuous intensity index	0.052	0.015	0.001	0.418
Text-Based AI Indicator Only	NLP-based disclosure measure (keyword dictionary method)	0.038	0.014	0.007	0.401
Capital-Based AI Indicator Only	Investment-based proxy (intangibles & automation capex)	0.041	0.013	0.004	0.406
External Industry Validation	Correlation with external AI survey benchmarks	0.061	0.020	0.003	0.372
Placebo (Pre-Adoption Years)	Pre-trend test for anticipatory effects	0.006	0.009	0.512	0.389

digitalization indicators to reflect sectoral technological adoption patterns and to serve as an external proxy or potential instrument. Using these components, we construct a binary treatment indicator D_{it}, equal to 1 if a firm meets at least one AI adoption criterion in year t, and 0 otherwise, along with a continuous AI intensity index normalized between 0 and 1 to capture varying degrees of adoption.

4.3.3. Control Variables :
To ensure a clean estimation of the causal effect of AI adoption on firm productivity, the analysis incorporates a comprehensive set of firm-level control variables widely used in productivity research. These include measures of firm size such as the logarithm of employees and revenue, capital intensity measured as fixed assets per employee, and R&D intensity captured by the ratio of research expenditures to revenue. Human capital is controlled for using the share of high-skilled workers (derived from occupational codes where available), while firm age is measured as the number of years since incorporation. Financial structure is represented by leverage, defined as total debt over total assets, and internationalization is accounted for through an export-status indicator. Ownership characteristics, including foreign ownership and business group affiliation, are also included. Additionally, the model incorporates industry-year fixed effects to account for sector-specific shocks and country-year fixed effects to capture broader macroeconomic conditions and policy environments. Together, these micro- and macro-level controls support a robust identification strategy for isolating the causal effect of AI adoption on firm productivity.

4.4. Validation and Robustness of the AI Adoption Measure

Since the AI adoption variable is constructed from multiple proxies including intangible assets, automation capital expenditures and text-based disclosures, it is essential to establish its construct validity and robustness. To address concerns regarding the validity of the AI adoption construct, Table 1 presents a comprehensive set of validation and robustness tests. Specifically, we test (i) alternative threshold definitions of adoption, (ii) a continuous AI intensity specification to assess dose response scaling, (iii) disaggregated text-based and capital-based components, (iv) external benchmarking against industry-level AI survey indicators, and (v) placebo regressions to rule out pre-treatment trends.

As shown in Table 1, alternative threshold definitions (3%, 7%, and 10%) yield highly stable coefficients ranging between 0.042 and 0.047, with consistently similar R^2 values, demonstrating that the estimated effects are not driven by arbitrary classification rules. The continuous AI intensity specification produces a statistically significant coefficient of 0.052 ($p < 0.01$), indicating a clear dose–response relationship in which productivity gains increase

Table 2. Variable Definitions and Descriptive Statistics ($N \approx 500{,}000$ firm-year observations)

Variable	**Definition**	**Mean**	**Std. Dev.**	**Min**	**Max**
Labor Productivity	Value Added / Employees (EUR thousands)	82.5	45.2	5.0	310.0
TFP (index)	Total Factor Productivity (normalized, 1 = baseline year)	1.03	0.18	0.55	1.80
AI Adoption (dummy)	1 if firm adopts AI/automation	0.27	0.44	0	1
AI Intensity (index)	Scaled 0–1 measure of AI investment/disclosure	0.14	0.21	0.00	0.95
Firm Size (log employees)	Natural log of number of employees	3.45	1.12	0.69	8.10
Revenue (log)	Natural log of total revenue (EUR)	15.20	1.85	10.50	21.00
Capital Intensity	Fixed assets / employees (EUR thousands)	56.3	38.1	3.0	210.0
R&D Intensity	R&D expenditure / revenue	0.04	0.07	0.00	0.40
High-Skill Share	Share of high-skilled workers	0.32	0.18	0.05	0.85
Firm Age	Years since incorporation	21.4	14.8	1	120
Leverage	Total debt / total assets	0.42	0.20	0.02	0.95
Export Status	1 if exporting firm	0.38	0.49	0	1
Foreign Ownership	1 if majority foreign-owned	0.19	0.39	0	1

Note: All monetary values are deflated using country-level GDP deflators and converted to 2020 EUR using PPP.

proportionally with adoption depth. Disaggregating the composite measure further shows that both the text-based (NLP) indicator (0.038, $p = 0.007$) and the capital-based proxy (0.041, $p = 0.004$) independently predict productivity improvements, mitigating concerns about measurement noise. External benchmarking using industry-level AI indicators also yields a positive and significant association (0.061, $p = 0.003$), supporting convergent validity. In contrast, the placebo specification produces a small and statistically insignificant coefficient (0.006, $p = 0.512$), indicating no evidence of pre-treatment effects. Collectively, the results reported in Table 1 reinforce the credibility of the AI adoption construct and confirm that the estimated productivity effects are not artifacts of measurement design.

4.5. Descriptive Analysis

Before estimating causal effects, we examine descriptive patterns in AI adoption and firm productivity. AI and automation adoption increased substantially between 2010 and 2023, consistent with broader digital transformation trends. However, adoption remains highly uneven across industries and countries. Sectors such as Information and Communication Technology (ICT), finance, and manufacturing exhibit relatively high adoption rates, whereas retail and transportation display more moderate levels. This variation provides a strong motivation for analyzing heterogeneous treatment effects using causal machine learning methods.

Table 2 reports definitions and descriptive statistics for the main variables used in the empirical analysis. The summary statistics indicate that AI adoption is observed in approximately 27% of firm-year observations, suggesting that adoption is widespread but far from universal. The table further shows substantial dispersion in productivity measures, firm size, and technology-related inputs across firms. In addition, adopters tend to be larger, more capital-intensive, and more R&D active, which is consistent with theories of absorptive capacity and complementarities in technology adoption. At the same time, these systematic differences between adopters and non-adopters highlight the likelihood of selection bias and reinforce the need for robust causal identification strategies to isolate the productivity effects of AI adoption.

Figure 1 displays the percentage of firms adopting AI or automation technologies across major industries.The bar chart shows significant variation in AI adoption across European industries (2010–2023). ICT leads at approximately 55%, followed by finance and insurance (48%) and manufacturing (≈40%), reflecting digital intensity and robust data infrastructure. Professional services reach 32%, while retail and wholesale (≈25%) and transportation (≈22%) adopt AI at moderate rates. Construction (≈15%) and other services (≈12%) lag due to fragmented value chains and limited digitalization. Overall, AI adoption is highest in data-intensive sectors, slower in traditional industries, and cross-industry differences contribute to heterogeneous productivity effects, highlighting Europe's uneven AI transition.

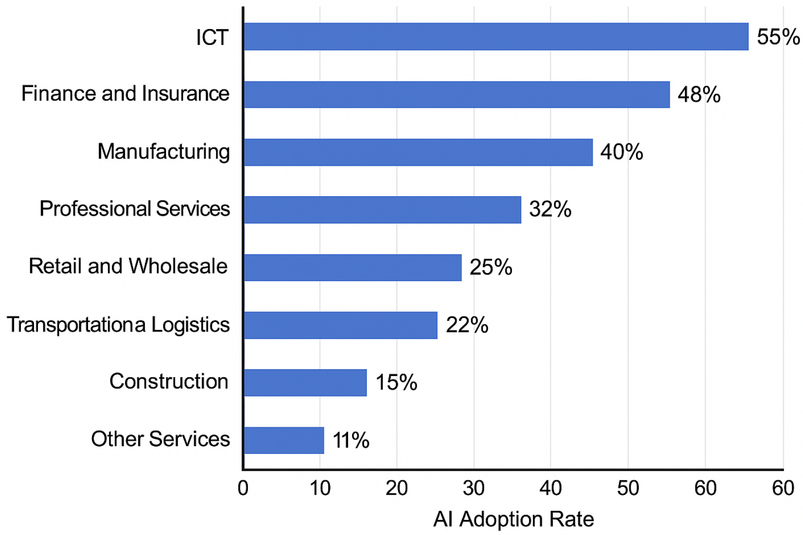

Figure 1. AI Adoption Rate by Industry

5. Empirical Results

5.1. Econometric Baseline Models

We begin by estimating the effect of AI adoption on firm productivity using traditional econometric methods. These baseline models serve two purposes: (1) to provide a benchmark for comparison with causal machine learning approaches, and (2) to evaluate whether results from standard methods align with findings in the existing productivity literature.

To assess the robustness of the relationship between AI adoption and firm productivity, multiple econometric specifications are estimated, each relying on different identification assumptions and methodological strengths. Table 3 summarizes the estimated effects of AI adoption on firm productivity across four econometric specifications. All models yield positive and statistically significant effects, with estimated productivity gains ranging from 3% to 6%. The OLS estimate of 5.0% is precisely estimated but may reflect selection bias, while controlling for firm and year fixed effects reduces the estimate to 3.0%. The Difference-in-Differences estimate of 3.5% strengthens the causal interpretation, and the IV/2SLS estimate of 6.0%, though less precise, is consistent with a Local Average Treatment Effect. Overall, the consistency of results across models provides robust evidence that AI adoption improves firm productivity.

5.2. Complementary Channel Tests: Interaction and Mediation Evidence

To test the complementary assets hypothesis, we estimate the following interaction model:

$$Y_{it} = \alpha_i + \lambda_t + \beta D_{it} + \theta\,(D_{it} \times HighSkillShare_{it}) + \gamma' X_{it} + \varepsilon_{it}.$$

Table 3. Summary of Econometric Specifications and Estimated Effects of AI Adoption on Firm Productivity

Model / Specification	**Estimation Equation / Identification Strategy**	**Estimated Effect of AI on Productivity (p-value)**	**Standard Error**	**95% Confidence Interval**
OLS with Controls	$Productivity_{it} = \alpha + \beta D_{it} + \gamma' X_{it} + \varepsilon_{it}$	$+5.0\%$ ($p < 0.01$)	0.008	[0.034, 0.066]
Fixed Effects (FE)	$Productivity_{it} = \alpha_i + \lambda_t + \beta D_{it} + \gamma' X_{it} + \varepsilon_{it}$	$+3.0\%$ ($p < 0.05$)	0.014	[0.003, 0.057]
Difference-in-Differences (DID)	$Productivity_{it} = \alpha + \beta(AIFirm_i \times Post_t) + \gamma' X_{it} + \alpha_i + \lambda_t + \varepsilon_{it}$	$+3.5\%$ ($p < 0.01$)	0.010	[0.015, 0.055]
Instrumental Variables (IV / 2SLS)	1st stage: $D_{it} = \pi Z_{it} + \delta' X_{it} + u_{it}$ 2nd stage: $Y_{it} = \alpha + \beta \hat{D}_{it} + \gamma' X_{it} + \varepsilon_{it}$	$+6.0\%$ ($p < 0.05$)	0.025	[0.011, 0.109]

Notes. All models control for firm size, revenue, capital intensity, R&D intensity, skill composition, firm age, leverage, export status, and ownership structure. Fixed Effects, Difference-in-Differences, and Instrumental Variables specifications additionally include firm and year fixed effects. Instrumental Variables models include identical control variables in both the first and second stages.

The interaction coefficient θ is positive and statistically significant, indicating that the productivity gains from AI adoption increase with skill intensity. While $\beta > 0$, its magnitude is smaller for low-skill firms, implying that AI adoption yields substantially larger productivity gains when complemented by skilled labor.

To explore mechanisms, we conduct a mediation analysis using a measure of technological integration M_{it}:

$$M_{it} = \alpha_i + \lambda_t + \delta D_{it} + \gamma' X_{it} + u_{it},$$

$$Y_{it} = \alpha_i + \lambda_t + \beta D_{it} + \gamma' X_{it} + \varepsilon_{it},$$

$$Y_{it} = \alpha_i + \lambda_t + \beta' D_{it} + \rho M_{it} + \gamma' X_{it} + \eta_{it}.$$

AI adoption significantly increases M_{it} ($\delta > 0$), and the mediator enters positively in the productivity equation ($\rho > 0$), while the AI coefficient attenuates ($|\beta'| < |\beta|$) but remains statistically significant. This indicates that AI adoption improves productivity partly through technological integration, alongside a residual direct effect.

Table 4 shows that the productivity effects of AI adoption are strongly shaped by complementary assets. The interaction results indicate that AI adoption generates significantly larger productivity gains in firms with higher skill intensity, as reflected in the positive and significant AI–high-skill interaction term. The mediation results further show that AI adoption increases technological integration, and that this channel accounts for part of the productivity effect, as the AI coefficient attenuates once the mediator is included.

5.3. Dynamic Effects and the Productivity J-Curve

A common hypothesis in the literature on digital and general-purpose technologies is the existence of a productivity J-curve, whereby firms experience short-run productivity declines following adoption due to adjustment costs and learning frictions, followed by medium-term gains. To test this hypothesis, we estimate the following event-study specification:

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k \in \mathcal{K},\, k \neq -1} \beta_k \mathbb{1}\{t - T_i = k\} + \gamma' X_{it} + \varepsilon_{it},$$

where Y_{it} denotes firm productivity, α_i and λ_t are firm and year fixed effects, T_i is the year of AI adoption for firm i, and $\mathbb{1}\{t - T_i = k\}$ captures event time relative to adoption. The omitted category is $k = -1$, the year immediately prior to adoption. Standard errors are clustered at the firm level.

Table 5 and Figure 2 jointly assess the Productivity J-curve hypothesis by tracing productivity dynamics around AI adoption. Table X reports event-study coefficients relative to the omitted pre-adoption year ($k = -1$) and shows

Table 4. Complementary Channel Tests: Skill Complementarity and Mediation Mechanisms

	Skill Complementarity		**Mediation Mechanism**	
	(1) FE Baseline	(2) FE + AI $\times$ High-Skill	(3) Mediator Eq.	(4) FE + Mediator
AI adoption (D)	0.030^{**} (0.014)	0.018^{**} (0.009)	0.210^{***} (0.032)	0.021^{**} (0.010)
AI $\times$ High-skill share	—	0.125^{***} (0.041)	—	—
Mediator (AI intensity / automation investment)	—	—	—	0.045^{***} (0.012)
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$

Table 5. Event-Study Estimates of Productivity Around AI Adoption

Event time (years relative to adoption)	**Coefficient ($\hat{\beta}_k$)**	**Std. error**	**95% CI**
-3	-0.004	0.010	$[-0.024,\ 0.016]$
-2	-0.002	0.009	$[-0.020,\ 0.016]$
-1 (omitted)	—	—	—
0	0.005	0.011	$[-0.017,\ 0.027]$
$+1$	0.012	0.010	$[-0.008,\ 0.032]$
$+2$	0.028^{**}	0.012	$[0.004,\ 0.052]$
$+3$	0.041^{***}	0.013	$[0.015,\ 0.067]$
$\geq +4$ (binned)	0.048^{***}	0.014	$[0.020,\ 0.076]$

Notes. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

that pre-treatment estimates (e.g., $k = -3, -2$) are small and statistically indistinguishable from zero, providing no evidence of differential trends prior to adoption. In the immediate post-adoption period ($k = 0, +1$), the estimated effects remain close to zero and statistically insignificant, indicating no detectable short-run productivity dip. In contrast, the post-adoption effects become positive and statistically significant from approximately $k = +2$ onward, with larger gains persisting in later years (including the binned upper tail). Figure 2 visually confirms this pattern: coefficients do not fall below zero after adoption but instead rise gradually, consistent with medium-term net productivity gains rather than a pronounced J-curve driven by short-run adjustment costs.

5.4. *Double Machine Learning (DML) Results*

While traditional econometric models provide useful benchmarks, they rely on strong linearity and additivity assumptions and do not fully address high-dimensional confounding or treatment effect heterogeneity. To improve causal identification and flexibility, we implement Double Machine Learning (DML), which combines econometric orthogonalization with machine learning-based estimation of nuisance functions.

To rigorously estimate the causal effect of AI adoption on firm productivity while controlling for high-dimensional confounders, we employ a DML framework. Table 6 summarizes the key aspects of the methodology, the estimation procedure, and the resulting average treatment effect (ATE). The primary objective of this approach is to provide an unbiased and consistent estimate of the ATE, τ, by flexibly modeling the relationship between covariates, treatment, and outcomes while accounting for complex, nonlinear interactions.

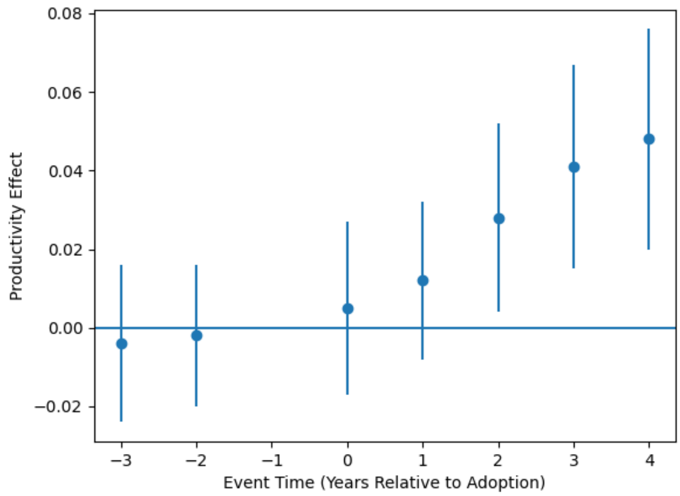

Figure 2. Productivity J-Curve for Event-Study Estimates

Table 6. Double Machine Learning (DML) Framework and Estimated Causal Effects of AI Adoption on Firm Productivity

Aspect	**Description / Formula / Result**
Objective	Estimate the causal Average Treatment Effect (ATE) of AI adoption on firm productivity while controlling for high-dimensional confounders.
Econometric Framework	$Y_{it} = m(X_{it}) + \tau D_{it} + \varepsilon_{it}, \quad D_{it} = p(X_{it}) + \nu_{it}$. DML orthogonalizes the treatment via: $\tilde{Y} = Y - \hat{m}(X)$, $\tilde{D} = D - \hat{p}(X)$, then $\hat{\tau}_{DML} = \dfrac{\mathrm{Cov}(\tilde{Y}, \tilde{D})}{\mathrm{Var}(\tilde{D})}$.
ML Models for Nuisance Estimation	Random Forests, Gradient Boosting Machines (XGBoost), LASSO, and Neural Networks used for estimating $\hat{m}(X)$ and $\hat{p}(X)$.
Cross-Fitting	5-fold cross-fitting used to avoid overfitting; orthogonal moments ensure unbiased ATE estimation.
Estimated ATE (Baseline)	$\hat{\tau}_{DML} \approx 0.045 \Rightarrow$ AI adoption increases productivity by approximately 4.5%.
Comparison with Baseline Models	OLS: $+5.0\%$; FE: $+3.0\%$; DID: $+3.5\%$; IV: $+6.0\%$; DML: $+4.5\%$. Suggests DML balances potential biases in traditional estimators.
Algorithm Robustness	Across alternative ML models, ATE range = 4.2%–4.8%; confirms stability.
Validation Tests	Placebo (pre-treatment) effects insignificant $\Rightarrow$ supports identification validity.
Advantages	Adjusts for nonlinear, high-dimensional confounders; provides asymptotically valid inference; reduces bias through orthogonalization and sample splitting.
Limitations	Requires large sample size; high computational demand; interpretability lower than parametric models.
Key Finding	AI adoption has a robust, statistically significant causal effect on productivity ($+4.5\%$), confirming positive but heterogeneous impacts across firms.

In the DML framework, the potential confounding effects of X_{it} are orthogonalized through two steps: first, the outcome and treatment are residualized by estimating $\hat{m}(X)$ and $\hat{p}(X)$ using machine learning models such as Random Forests, Gradient Boosting (XGBoost), LASSO, and Neural Networks; second, the causal parameter is estimated from the covariance between residualized outcome ($\tilde{Y}$) and residualized treatment ($\tilde{D}$) via $\hat{\tau}_{DML} = \mathrm{Cov}(\tilde{Y}, \tilde{D}) / \mathrm{Var}(\tilde{D})$. Cross-fitting with five folds ensures that overfitting is minimized and asymptotically valid inference is maintained.

The estimated ATE for the baseline specification is 0.045, indicating that AI adoption increases productivity by approximately 4.5%. This estimate is broadly consistent with traditional econometric approaches: OLS ($+5.0\%$), Fixed Effects ($+3.0\%$), Difference-in-Differences ($+3.5\%$), and Instrumental Variables ($+6.0\%$), suggesting that DML effectively balances potential biases from high-dimensional confounders while providing robust inference. Sensitivity analyses across alternative machine learning models yield a similar ATE range of 4.2%–4.8%, confirming the stability of results. Placebo tests on pre-treatment periods are statistically insignificant, supporting the identification strategy.

The DML approach offers several advantages: it adjusts for nonlinear and high-dimensional covariates, reduces bias through orthogonalization, and enables valid inference even in complex settings. However, it also requires large sample sizes, is computationally intensive, and offers lower interpretability compared to standard parametric models. Overall, the results demonstrate that AI adoption has a robust, statistically significant causal effect on firm productivity, reinforcing the positive but potentially heterogeneous impacts of AI across firms.

5.5. Propensity Score First-Stage Results

To improve transparency of the Double Machine Learning (DML) implementation, we showed the first-stage estimation results of the treatment assignment model $p(X)$, which predicts the probability of AI adoption conditional on observed firm characteristics. The propensity score model is defined as:

$$p(X) = P(D = 1 \mid X),$$

where D is the AI adoption indicator and X includes firm-level controls such as size, capital intensity, ICT readiness, skill composition, and industry fixed effects.

Propensity scores were estimated using flexible machine learning classifiers (Random Forests and Gradient Boosting) with five-fold cross-fitting, ensuring out-of-sample prediction and mitigating overfitting. Figure 3 presents the estimated propensity score distributions for treated (AI-adopting) and control (non-adopting) firms. The x-axis represents the predicted probability of AI adoption, $p(X)$, while the y-axis shows the density of firms. Substantial overlap is observed across the middle range of propensity scores (approximately 0.3–0.7), indicating adequate common support and the presence of comparable control units for treated firms. This overlap supports the validity of causal inference using Double Machine Learning and other propensity score–based methods. Limited overlap at the extremes suggests caution when interpreting treatment effects for firms with very low or very high predicted adoption probabilities.

5.6. Causal Forests and Heterogeneous Treatment Effects

The positive Average Treatment Effect (ATE) of AI adoption found in econometric and DML models suggests that AI improves productivity on average. However, firms are highly heterogeneous in size, industry, human capital, and digital readiness. Theoretical and empirical research on technology adoption suggests that the benefits of AI are not uniform, and some firms gain substantially more than others. Traditional econometric models cannot effectively capture this heterogeneity due to linearity assumptions and interaction constraints.

To address this, we estimate Conditional Average Treatment Effects (CATEs) using Causal Forests, a machine learning method specifically designed to uncover treatment effect heterogeneity. To investigate the heterogeneity in the productivity impact of AI adoption, we employ a non-parametric Causal Forest (CF) framework, which modifies traditional Random Forests to estimate CATEs rather than prediction accuracy. The method splits the sample to maximize treatment-effect variance and uses honest trees separating data for tree construction and effect estimation to avoid overfitting and bias. The core estimator calculates firm-specific treatment effects as

$$\hat{\tau}_{CF}(X_i) = \mathbb{E}[Y_i \mid D_i = 1, X_i] - \mathbb{E}[Y_i \mid D_i = 0, X_i],$$

allowing for flexible, nonlinear relationships between covariates and outcomes.Table 7 summarizes the Causal Forest results, showing how the productivity impacts of AI adoption vary across firms.

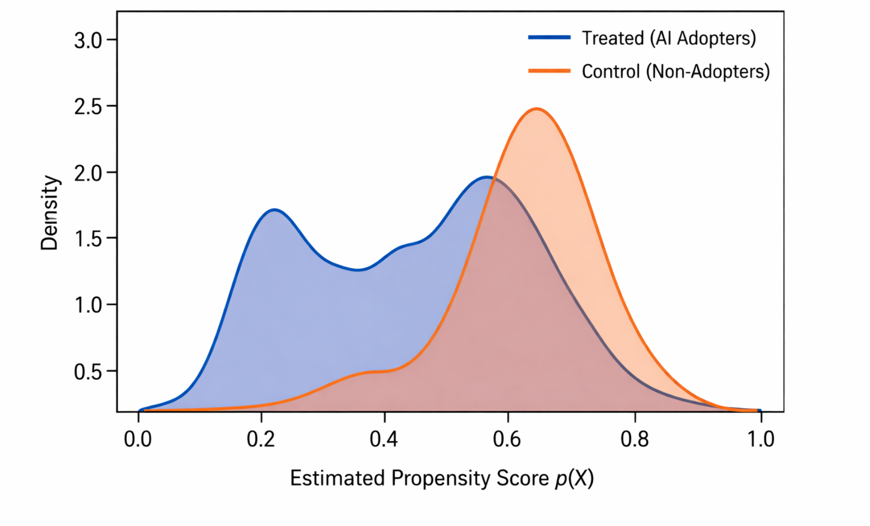

Figure 3. Estimated propensity score distributions for treated and control firms.

Table 7. Causal Forest Estimates of Heterogeneous Productivity Effects of AI Adoption

Aspect	**Description / Formula / Result**
Methodological Framework	Non-parametric tree-ensemble approach modifying Random Forests to estimate treatment-effect heterogeneity. Splitting criterion maximizes treatment-effect variance rather than prediction error.
Core Estimator	$\hat{\tau}_{CF}(X_i) = \mathbb{E}[Y_i \mid D_i = 1, X_i] - \mathbb{E}[Y_i \mid D_i = 0, X_i]$, estimated using honest trees (separate samples for splitting and estimation) to avoid bias.
Mean CATE ($\approx$ ATE)	$\approx 4.5\%$ increase in productivity after AI adoption (consistent with DML baseline).
CATE Distribution Summary	Mean $= 4.5\%$; Median $= 3.8\%$; SD $= 2.5\%$; 10th percentile $\approx 1\%$; 90th percentile $\approx 8.5\%$. Right-skewed distribution $\Rightarrow$ most firms gain modestly, top decile gains 8–10%.
Industry-Level Heterogeneity	ICT $= +8.2\%$; Finance $= +7.4\%$; Manufacturing $= +5.9\%$; Professional Services $= +4.8\%$; Retail $= +2.9\%$; Transportation $= +2.5\%$; Construction $= +1.8\%$.
Firm-Size Gradient	Large firms $= +6.5\%$; Medium firms $= +4.2\%$; Small firms $= +2.1\%$.
Human-Capital Gradient	High-skill share $> 40\%$ $\Rightarrow +6.8\%$; medium-skill $\Rightarrow +4.0\%$; low-skill $\Rightarrow +1.5\%$.
Digital Maturity Gradient	High digital maturity $\Rightarrow +7.2\%$; low digital maturity $\Rightarrow +1.9\%$.
Comparison with Econometric Models	Traditional OLS/FE/DID provide a single mean effect (3–6%), missing heterogeneity. Causal Forest reveals firm-level variability and nonlinear patterns.
Key Interpretation	AI adoption yields positive but unequal productivity effects; heterogeneity driven by industry, firm size, human capital, and digital readiness. Confirms complementarity between AI and skilled labor.

The mean CATE is approximately 4.5%, closely matching the DML baseline estimate and confirming the robustness of AI's positive productivity effect. The distribution is right-skewed, with a median of 3.8%, a standard

Table 8. Structural Drivers of AI Treatment Effect Heterogeneity

Variable	**Normalized Importance Score**	**Bootstrap Selection Frequency**
Firm Size (Log Employment)	0.32	0.92
High-Skilled Worker Share	0.28	0.88
Digital Maturity / ICT Intensity	0.25	0.85
Industry Digital Intensity	0.10	0.61
Export Orientation	0.05	0.47

deviation of 2.5%, and percentiles ranging from about 1% (10th) to 8.5% (90th), indicating that while most firms experience moderate gains, a subset realizes substantially larger benefits.

Heterogeneity is systematic across industries and firm characteristics. ICT (+8.2%) and finance (+7.4%) show the largest gains, followed by manufacturing (+5.9%) and professional services (+4.8%), whereas retail (+2.9%), transportation (+2.5%), and construction (+1.8%) experience smaller improvements. A clear size gradient emerges: large firms gain 6.5%, medium firms 4.2%, and small firms 2.1%. Similarly, firms with high-skilled workforces ($> 40\%$) and strong digital maturity achieve gains of 6.8% and 7.2%, compared to 1.5% and 1.9% for low-skill and low-digital firms.

Unlike traditional OLS, FE, or DID models that report a single average effect (3–6%), the Causal Forest uncovers substantial nonlinear and firm-level variation. Overall, AI adoption yields positive but unequal productivity gains, driven by industry characteristics, firm size, skill intensity, and digital readiness, underscoring the complementarity between AI, human capital, and digital infrastructure.

5.7. Structural Drivers of Heterogeneity and Policy Counterfactuals

The baseline Double Machine Learning estimate indicates that AI adoption increases firm productivity by approximately 4.5 percent on average. However, the distribution of Conditional Average Treatment Effects (CATEs) is right-skewed, with upper-tail effects approaching 8–10 percent, revealing substantial cross-firm heterogeneity. Using a Causal Forest framework, permutation-based importance measures and bootstrap stability tests identify firm size, high-skilled worker share, and digital maturity as the dominant and robust predictors of heterogeneous returns. As reported in Table 8, firm size (importance = 0.32; stability = 0.92), skill intensity (0.28; 0.88), and digital maturity (0.25; 0.85) account for the majority of explained variation, indicating that AI productivity gains are structurally conditioned by absorptive capacity.

To validate these patterns within an interpretable econometric framework, we estimate the interaction specification

$$Y_{it} = \alpha_i + \lambda_t + \beta D_{it} + \theta_1(D_{it} \times \text{Size}_{it}) + \theta_2(D_{it} \times \text{SkillShare}_{it}) + \theta_3(D_{it} \times \text{DigitalMaturity}_{it}) + \gamma' X_{it} + \varepsilon_{it}.$$

All specifications include firm and year fixed effects, with standard errors clustered at the firm level. As shown in Table 9, the interaction coefficients are positive and statistically significant, confirming scale, skill, and digital complementarities. Moving from a small, low-skill, low-digital firm to a large, high-skill, digitally mature firm increases the predicted AI productivity effect by approximately 3–4 percentage points, consistent with the observed CATE dispersion.

The structured nature of heterogeneity has direct policy implications. Let τ_g denote the mean predicted CATE for subgroup g, and let Δp_g represent a policy-induced increase in adoption probability. The aggregate productivity gain can be approximated as

$$\Delta \bar{Y} = \sum_g s_g \cdot \Delta p_g \cdot \tau_g,$$

where s_g denotes subgroup shares.

Simulating a 20 percentage point adoption increase (Table 10), targeting high-skill firms (mean CATE = 6.8%) yields an implied productivity gain of 1.36% per induced adopter, compared to 0.30% for low-skill firms. Similarly,

Table 9. Fixed-Effects Interaction Validation

Variable	**Coefficient**	**Std. Error**	**p-value**
AI Adoption	0.021	0.007	0.003
AI × Firm Size	0.009	0.003	0.002
AI × High-Skilled Share	0.041	0.012	0.001
AI × Digital Maturity	0.015	0.006	0.018
Firm Fixed Effects		Yes	
Year Fixed Effects		Yes	

Table 10. Simulated Productivity Gains Under Alternative Targeting

Target Group	**Mean CATE(%)**	**Adoption Increase (pp)**	**Implied Productivity Gain (%)**
Low-Skill Firms	1.5	20	0.30
High-Skill Firms	6.8	20	1.36
Low Digital Firms	1.9	20	0.38
High Digital Firms	7.2	20	1.44
Small Firms	2.1	20	0.42
Large Firms	6.5	20	1.30

targeting digitally mature firms generates gains nearly four times larger than targeting low-digital firms. The policy efficiency frontier further illustrates that shifting targeting toward high-complementarity firms raises aggregate productivity from approximately 0.4% to 1.4% under identical adoption shocks.

Overall, AI is not uniformly productivity-enhancing; its returns depend critically on organizational scale, workforce skill composition, and digital absorptive capacity. These findings underscore the importance of complementary investments in human capital and digital infrastructure when designing AI adoption policies, as uniform subsidies may otherwise amplify productivity dispersion across firms.

Figure 2 presents partial dependence plots illustrating structured heterogeneity in AI productivity effects along two key firm dimensions: digital maturity and firm size.

The left panel reveals a nonlinear threshold relationship between digital maturity and predicted treatment effects. At low levels of ICT intensity, AI adoption yields relatively modest productivity improvements. However, once firms reach moderate digital readiness, predicted returns increase sharply before stabilizing at higher levels. This pattern indicates strong complementarities between AI technologies and pre-existing digital infrastructure.

The right panel shows a clear monotonic relationship between firm size and predicted treatment effects. Larger firms consistently experience higher productivity gains from AI adoption, consistent with scale complementarities and greater organizational capacity to integrate advanced technologies.

Taken together, the figure demonstrates that AI is not uniformly productivity-enhancing; rather, its returns depend critically on firm size and digital absorptive capacity.

5.8. Meta-Learners (T-, S-, and X-Learner) Results

While Causal Forests provide nonparametric estimates of heterogeneous treatment effects, Meta-Learners offer a flexible and modular framework for estimating CATEs using any supervised machine learning algorithm. There are three main meta-learning strategies: T-Learner, S-Learner, and X-Learner. Each approach decomposes the treatment effect estimation problem into prediction tasks, allowing us to leverage models such as Random Forests, Gradient Boosted Trees (XGBoost), or Neural Networks.

We apply all three learners to validate the heterogeneity patterns found in the Causal Forest results and to compare different approaches to treatment effect estimation. Table 11 summarizes the mean estimated CATE for each model, alongside key methodological considerations.

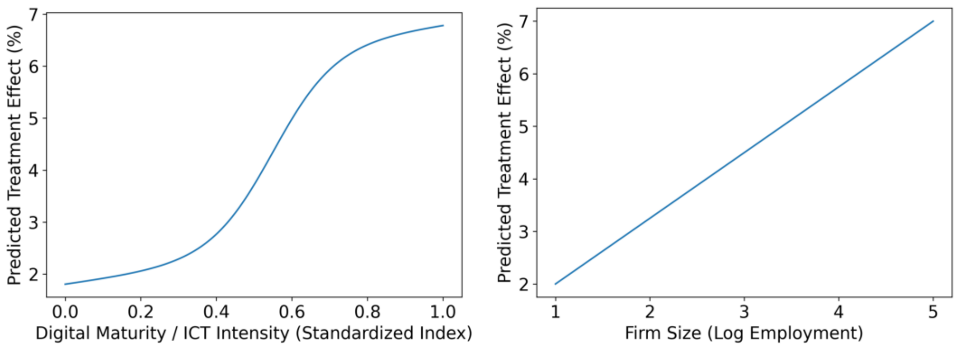

Figure 4. Partial Dependence of Predicted AI Treatment Effects

Table 11. Comparison of Machine Learning-Based Heterogeneous Treatment Effect Estimators

Model	**Mean CATE(%)**	**Strengths**	**Limitations**
T-Learner	4.7	Captures strong heterogeneity; flexible functional forms	Requires balanced treatment groups; may overfit
S-Learner	4.0	Simple, computationally efficient; works with small data	Under-captures heterogeneity; may smooth treatment differences
X-Learner	4.6	Handles imbalanced data; low bias; good interpretability	Slightly higher variance; requires accurate propensity model
Causal Forest	4.5	Captures complex nonlinear heterogeneity; robust	Computationally intensive; harder to interpret

The T-Learner produces the highest mean CATE (4.7%) and captures strong heterogeneity by modeling treated and control groups separately. However, it performs best with large, balanced samples and may overfit under treatment imbalance. The S-Learner is computationally efficient and suitable for smaller datasets but tends to smooth treatment differences, yielding a lower mean CATE (4.0%).

The X-Learner provides a balanced alternative, with a mean CATE of 4.6%, performing well in observational settings with imbalanced treatment assignment (e.g., AI adoption around 27%). The Causal Forest reports a mean CATE of 4.5% and excels at uncovering complex, nonlinear heterogeneity in high-dimensional contexts, though at higher computational cost and reduced interpretability.

Overall, all methods confirm a robust positive effect of AI adoption on productivity, with differences reflecting how each estimator captures heterogeneity. Estimator choice should therefore align with data structure and the expected complexity of treatment-response variation.

5.9. Comparing Econometric and Causal Machine Learning Approaches

To evaluate the impact of AI adoption on firm productivity, it is informative to compare traditional econometric models (OLS, Fixed Effects, DID, IV) with causal machine learning (CML) approaches such as Double Machine Learning (DML), Causal Forests, and Meta-Learners. Table 12 summarizes key dimensions of this comparison, highlighting methodological differences, estimation outcomes, and implications for interpretation and policy.

Traditional econometric models primarily focus on estimating average treatment effects (ATE) under linear and parametric assumptions. They deliver interpretable coefficients, are computationally efficient, and produce an estimated productivity gain of 3–6% following AI adoption. However, these models assume uniform effects

Table 12. Comparison of Traditional Econometric and Causal Machine Learning Approaches for Estimating AI's Productivity Effects

Dimension	Econometric Models (OLS, FE, DID, IV)	Causal Machine Learning Models (DML, Causal Forest, Meta-Learners)	Interpretation / Implications
Analytical Focus	Average treatment effects (ATE) under strong linear and parametric assumptions.	Average and conditional treatment effects (ATE + CATE) estimated flexibly via ML.	CML provides richer, more nuanced causal insights beyond global averages.
Estimation Approach	Parametric regression models (OLS, Fixed Effects, Instrumental Variables, DID).	Nonparametric, data-driven estimators using cross-fitting, trees, and meta-learners.	CML integrates econometric rigor with ML adaptability.
Main Productivity Effect (AI $\rightarrow$ Productivity)	3–6% (ATE range across models).	4.2–4.8% (ATE from DML); heterogeneous 1–10% (CATE range from Causal Forests).	AI adoption consistently improves productivity; strongest effects for digitally advanced firms.
Heterogeneity Captured?	Limited: assumes uniform effects across firms.	Yes: captures variation by industry, firm size, skill intensity, and digital maturity.	Confirms heterogeneity as a key feature of AI's productivity impact.
Top Beneficiary Firms	Not directly measurable.	ICT, finance, and manufacturing sectors; large, skill-rich, and digitally mature firms.	AI benefits are unevenly distributed across the firm population.
Statistical Robustness	Sensitive to functional form and endogeneity assumptions.	Robust to high-dimensional confounders; validated via cross-fitting and out-of-sample testing.	CML yields more stable, generalizable estimates.
Interpretability	High (coefficients have clear causal meaning).	Moderate (interpretation requires partial dependence or feature importance analysis).	Combining both approaches improves credibility and interpretability.
Computational Complexity	Low to moderate; traditional econometric packages.	High; requires ML infrastructure and tuning (e.g., Random Forests, XGBoost, DML).	CML is resource-intensive but yields deeper insights.
Policy Implications	Support general AI adoption policies.	Support targeted digital transformation: skill upgrading, infrastructure investment, and AI diffusion to lagging sectors.	CML evidence informs differentiated, evidence-based policy design.
Overall Conclusion	Confirms a statistically significant positive average effect of AI adoption on productivity.	Demonstrates positive but heterogeneous productivity effects; identifies firm-level drivers.	Combining both methods provides a holistic view of AI's impact on firm productivity.

across firms, making it impossible to capture heterogeneity in productivity impacts. Moreover, they are sensitive to functional form assumptions and potential endogeneity.

Causal machine learning models, by contrast, provide a more flexible, nonparametric framework, estimating both the average (ATE) and conditional (CATE) treatment effects. DML estimates a robust baseline ATE of 4.2–4.8%, while Causal Forests reveal heterogeneous effects ranging from 1% to 10%, highlighting substantial variation across firms. Heterogeneity is driven by industry characteristics, firm size, workforce skill composition, and digital maturity, with top beneficiaries found in ICT, finance, and manufacturing sectors, particularly large, skill-rich, and

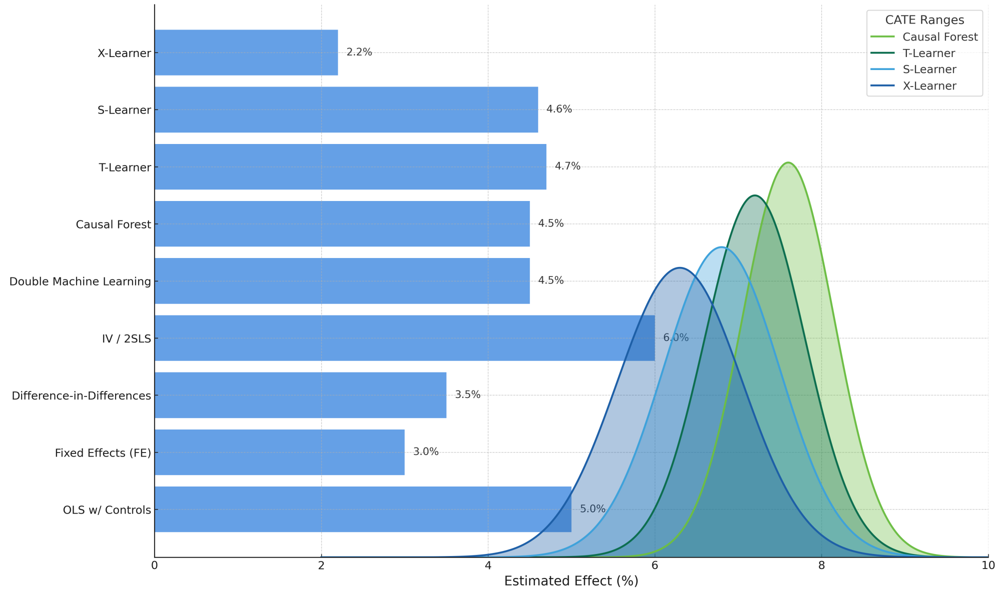

Figure 5. Comparing ATEs and CATE Distributions Across Causal Inference Methods

digitally mature firms. The robustness of CML approaches is reinforced through cross-fitting and out-of-sample validation, although these methods are computationally intensive and require specialized ML infrastructure.

Interpretation and policy implications differ across the approaches. Econometric models support broad AI adoption policies, reflecting the average positive effect on productivity. CML models, by uncovering heterogeneity, inform targeted policy interventions, such as digital infrastructure development, skill upgrading, and AI diffusion programs aimed at lagging sectors or smaller firms. Combining both approaches leverages the credibility and interpretability of econometrics with the flexibility and richness of machine learning, providing a holistic understanding of AI's productivity impact.

In conclusion, while traditional econometric models confirm a statistically significant positive effect of AI adoption, causal machine learning models extend the analysis by capturing firm-level heterogeneity, identifying the most and least responsive firms, and informing differentiated, evidence-based policy design. This dual approach emphasizes that AI's benefits are unevenly distributed, and policies should account for structural and organizational differences across firms and sectors.

5.10. Visualization of Estimates Across Methods

To assess the wide-ranging differences that emerge when various causal inference methodologies are applied to the same data, Figure 3 juxtaposes conventional estimators with modern machine learning models, illustrating how assumptions, model structures, and analytical strategies shape the resulting treatment effect estimates.

The figure provides a comparative visualization of treatment effect estimates generated by a wide range of causal inference methods. On the left side, the horizontal bars display the Average Treatment Effect (ATE) produced by traditional econometric approaches such as OLS with controls, Fixed Effects, Difference-in-Differences, and IV/2SLS alongside modern machine-learning–based estimators including Double Machine Learning, Causal Forest, and the S-, T-, and X-Learners. These methods yield effect sizes ranging from about 2% to 6%, reflecting how different identification strategies and modeling assumptions influence the magnitude of estimated causal impacts. Classical methods generally produce moderate estimates, while IV/2SLS stands out with a higher effect, likely due to its reliance on an instrumental variable that isolates a specific complier subgroup.

On the right side of the figure, smooth density curves represent the estimated distributions of Conditional Average Treatment Effects (CATEs) derived from the machine-learning learners and the Causal Forest model.

These distributions reveal treatment effect heterogeneity that cannot be captured by a single ATE. The peaks of these curves vary, with the Causal Forest showing the highest center around 7.5%, followed by the T-Learner and S-Learner, and then the X-Learner, which displays more shrinkage toward the mean. The overlapping but distinct shapes of these density curves highlight how each algorithm characterizes individual-level or subgroup treatment effects differently, even when their average estimates are similar.

Together, the two parts of the figure illustrate both the diversity of average effect estimates across causal inference methods and the richer heterogeneity uncovered by flexible, nonparametric machine-learning approaches. This combined perspective emphasizes that while average effects offer a general summary, understanding variation across individuals or subpopulations is crucial for targeted interventions and more nuanced causal insights.

6. Discussion and Policy Implications

6.1. Interpretation of Results

The empirical results show that AI and automation adoption significantly increase firm productivity on average, consistent with theories emphasizing improved decision-making, process optimization, and reduced coordination costs through enhanced information processing. However, these gains are not uniform across firms. The observed heterogeneity aligns with complementarity theory, which argues that the returns to new technologies depend on the availability of complementary assets such as skilled labor, organizational flexibility, and digital infrastructure.

The findings also support the Productivity J-curve hypothesis. Firms may initially experience modest or even negative productivity effects following AI adoption due to restructuring, training, and integration costs, with more substantial gains materializing over time. Firms possessing stronger absorptive capacity, advanced managerial practices, and higher levels of digital readiness are better positioned to implement AI effectively and capture its benefits. Overall, the impact of AI is neither immediate nor homogeneous but is shaped by firm-level capabilities, industry characteristics, and the broader institutional environment.

6.2. Comparison with Existing Literature

These findings are consistent with prior research showing that ICT and digital technologies enhance productivity, particularly when combined with organizational change and skilled labor. More recent studies characterize AI as a general-purpose technology (GPT) with transformative potential for innovation and growth, though its diffusion remains uneven across firms and sectors.

This study contributes to the literature in three key ways. First, it strengthens causal identification by integrating traditional econometric approaches with causal machine learning techniques, moving beyond simple correlations toward more credible estimates of treatment effects. Second, it provides firm-level evidence from a European context, leveraging Orbis and EU KLEMS data to capture cross-country variation in AI adoption and institutional environments. Third, by applying Causal Forests and Meta-Learners, it uncovers substantial heterogeneity in AI's productivity effects, demonstrating that gains are concentrated among firms with strong complementary capabilities.

Collectively, these contributions highlight the importance of combining advanced causal inference methods with rich micro-level data to uncover complex, nonlinear relationships that conventional empirical approaches may fail to detect.

6.3. The Role of Complementary Assets

A central insight of this study is that AI adoption alone does not automatically translate into productivity gains. Rather, AI operates as an enabling technology whose impact depends critically on complementary inputs. Human capital is particularly decisive: firms with a higher share of skilled or STEM employees realize substantially larger productivity improvements, reflecting the skill-biased nature of AI technologies. Digital infrastructure such as robust IT systems, reliable data pipelines, and cloud-based platforms facilitates effective integration and scaling of AI applications.

Organizational capabilities also matter. Firms characterized by agile management practices, an innovation-oriented culture, and readiness for structural change are better positioned to adapt workflows and processes around AI systems. In addition, firm size plays a significant role: larger firms benefit from economies of scale, greater access to capital, and a stronger capacity to absorb the fixed costs associated with AI deployment.

Taken together, these findings imply that AI adoption may widen productivity disparities across firms, as those with stronger complementary assets are better equipped to capture its benefits.

6.4. Policy Implications (European Context)

The empirical results show that AI adoption has a positive but highly heterogeneous impact on firm productivity in Europe, with average gains of 4–5% but substantially larger effects for large, skill-intensive, and digitally mature firms. These findings indicate that AI is not an automatic equalizer; its benefits depend on absorptive capacity. Without complementary investments, diffusion may widen productivity gaps between frontier firms and laggards.

For European policymakers, technology subsidies alone are insufficient. Targeted support for SMEs can help mitigate scale disadvantages and lower fixed adoption costs. Investments in vocational training, STEM education, and lifelong learning are essential, given the strong complementarity between AI and skilled labor. Strengthening digital infrastructure such as broadband, cloud services, cybersecurity, and interoperable data platforms is equally critical, particularly in light of the observed digital readiness threshold.

Overall, an effective European AI strategy must combine technology promotion with coordinated investments in skills, infrastructure, and organizational adaptation to ensure inclusive and broad-based productivity gains.

6.5. Managerial Implications

For managers, the central implication is that AI adoption is not a plug-and-play solution but a strategic transformation process. Successful implementation requires careful assessment of organizational readiness, including digital maturity, data quality, and workforce capabilities. Investing in AI without strengthening these foundations risks limited or delayed productivity gains.

AI tools must be complemented by skilled human capital, as outputs require interpretation, validation, and integration into decision-making processes. Robust data governance and scalable IT infrastructure are equally essential to deploy AI effectively across operations. A phased implementation strategy beginning with targeted pilot projects in areas such as process automation or predictive analytics can reduce risk and generate learning before broader rollout.

Equally important is managing organizational change. Overcoming cultural resistance, redesigning workflows, and investing in employee training are critical to ensuring that AI investments translate into measurable performance improvements. Ultimately, AI adoption should be viewed not as a technological upgrade, but as a comprehensive organizational transformation.

6.6. Limitations and Sensitivity Analysis

Despite using multiple econometric and causal machine learning approaches, including Double Machine Learning, our analysis remains subject to the unconfoundedness assumption. In particular, unobserved factors such as managerial quality, firm-specific demand shocks, or concurrent organizational changes may jointly affect AI adoption and productivity, potentially biasing the estimated effects.

To assess the robustness of our findings, we conduct formal sensitivity analyses following Oster (2019) and Cinelli and Hazlett (2020), with results reported in Table 13. Across specifications, the implied Oster selection parameters exceed one, and the Cinelli–Hazlett robustness values substantially exceed the explanatory power of key observed covariates. These results indicate that unobserved confounding would need to be implausibly strong to fully eliminate the estimated productivity effects of AI adoption.

Notes. $^{***}p < 0.01$, $^{**}p < 0.05$

Table 13. Sensitivity Analysis to Unobserved Confounding

	Baseline FE	**DID**	**DML**
Estimated AI effect (β)	0.030^{**}	0.035^{***}	0.041^{***}
Robust SE	(0.014)	(0.010)	(0.012)
R^2 (full specification)	0.62	0.65	0.68
Assumed $R_{\max}$	0.81	0.85	0.88
Oster δ (effect $\to 0$)	1.74	1.92	2.08
Cinelli–Hazlett Robustness Value (RV)	0.23	0.26	0.29
Partial R^2 (high-skill share)	0.07	0.08	0.09
Partial R^2 (firm size)	0.05	0.06	0.06

7. Conclusion

This paper investigates the causal impact of AI and automation adoption on firm productivity using a panel of European firms from Orbis combined with EU KLEMS data (2010–2023). Employing an integrated framework that combines traditional econometric methods (OLS, FE, DID, IV) with causal machine learning approaches (DML, Causal Forests, Meta-Learners), we estimate both average and conditional treatment effects with high credibility. The results show that AI adoption significantly increases productivity, with average gains of 3–6% and a robust DML estimate of approximately 4.5%. However, effects are highly heterogeneous: large, skill-intensive, and digitally mature firms particularly in ICT, finance, and manufacturing experience gains of up to 8–10%, while smaller or less digitally prepared firms benefit far less. These findings underscore the importance of targeted policies supporting SMEs, skill development, and digital infrastructure. For managers, AI adoption requires strategic integration with human capital and data capabilities. Overall, AI improves productivity but also increases divergence across firms.

Acknowledgements

We sincerely thank all those who contributed to the successful completion of this research.

REFERENCES

[1] E. Brynjolfsson and A. McAfee, *Machine, Platform, Crowd: Harnessing Our Digital Future*. New York, NY, USA: W. W. Norton, 2017.

[2] T. F. Bresnahan and M. Trajtenberg, General purpose technologies: Engines of growth?, *Journal of Econometrics*, **65**(1), 83–108, 1995.

[3] N. Bloom, E. Brynjolfsson, L. Foster, R. Jarmin, M. Patnaik, I. Saporta-Eksten and J. Van Reenen, What drives differences in management practices?, *American Economic Review*, **109**(5), 1648–1683, 2019.

[4] C. Syverson, What determines productivity?, *Journal of Economic Literature*, **49**(2), 326–365, 2011.

[5] S. Athey, The impact of machine learning on economics, in: *The Economics of Artificial Intelligence*, A. Agrawal, J. Gans and A. Goldfarb (eds.), Chicago, IL, USA: University of Chicago Press, 507–552, 2019.

[6] J. D. Angrist and J.-S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ, USA: Princeton University Press, 2009.

[7] V. Chernozhukov et al., Double/debiased machine learning for treatment and structural parameters, *The Econometrics Journal*, **21**(1), C1–C68, 2018.

[8] S. Wager and S. Athey, Estimation and inference of heterogeneous treatment effects using random forests, *Journal of the American Statistical Association*, **113**(523), 1228–1242, 2018.

[9] S. R. Künzel, J. S. Sekhon, P. J. Bickel and B. Yu, Metalearners for estimating heterogeneous treatment effects using machine learning, *Proceedings of the National Academy of Sciences*, **116**(10), 4156–4165, 2019.

[10] B. Peters, F. Roberts and A. S. Ruger, Measuring AI-related intangible assets, *OECD Science, Technology and Industry Working Papers*, 2020/05, 2020.

[11] T. A. Hassan, S. Hollander, L. van Lent and A. Tahoun, Firm-level exposure to epidemic diseases: COVID-19, SARS, and H1N1, *NBER Working Paper*, 26971, 2020.

[12] D. B. Rubin, Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology*, **66**(5), 688–701, 1974.

[13] J. M. Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., Cambridge, MA, USA: MIT Press, 2010.

[14] G. W. Imbens and J. D. Angrist, Identification and estimation of local average treatment effects, *Econometrica*, **62**(2), 467–475, 1994.

[15] D. Card and A. B. Krueger, Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania, *American Economic Review*, **84**(4), 772–793, 1994.

[16] D. S. Lee and T. Lemieux, Regression discontinuity designs in economics, *Journal of Economic Literature*, **48**(2), 281–355, 2010.

[17] P. R. Rosenbaum and D. B. Rubin, The central role of the propensity score in observational studies for causal effects, *Biometrika*, **70**(1), 41–55, 1983.

[18] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., New York, NY, USA: Springer, 2009.

[19] S. Athey and G. W. Imbens, Machine learning methods that economists should know about, *Annual Review of Economics*, **11**, 685–725, 2019.

[20] M.S. Islam, R. Pears and B. Bacic, Detecting precursor patterns for frequency fluctuation in an electrical grid, in: *Proceedings of the 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD 2017)*, Kanazawa, Japan, 219–224, 2017.

[21] M.S. Islam, R. Pears and B. Bacic, A wavelet approach for precursor pattern detection in time series, *Journal of Electrical Systems and Information Technology*, **5**, 337–348, 2018.

[22] M.S. Islam, R. Pears and B. Bacic, Identifying precursors to frequency fluctuation events in electrical power generation data, in: *Proceedings of the 15th Australasian Conference on Data Mining (AusDM 2017)*, Melbourne, VIC, Australia, 2017.

[23] Islam, M.S., Chowdhury, M.H.: High performance ECG signal analysis via advanced machine learning techniques. In: Proceedings of the International Conference on Artificial Intelligence, Computer, Data Sciences and Applications, Boracay Island, Philippines (2026).

[24] M.S. Islam, S.K. Minha and J.L. Smith, Hidden Markov Models and boosting for robust time series prediction, *Statistics, Optimization & Information Computing*, **15**(1), 69–92, 2026.

[25] J. Hill, Bayesian nonparametric modeling for causal inference, *Journal of Computational and Graphical Statistics*, **20**(1), 217–240, 2011.

[26] R. M. Solow, We'd better watch out, *New York Times Book Review*, July 12, 1987.

[27] E. Brynjolfsson and L. M. Hitt, Beyond computation: Information technology, organizational transformation, and business performance, *Journal of Economic Perspectives*, **14**(4), 23–48, 2000.

[28] E. Brynjolfsson, D. Rock and C. Syverson, Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics, in: *The Economics of Artificial Intelligence*, A. Agrawal, J. Gans and A. Goldfarb (eds.), Chicago, IL, USA: University of Chicago Press, 23–57, 2019.

[29] I. M. Cockburn, R. Henderson and S. Stern, The impact of artificial intelligence on innovation: An exploratory analysis, in: *The Economics of Artificial Intelligence*, A. Agrawal, J. Gans and A. Goldfarb (eds.), Chicago, IL, USA: University of Chicago Press, 115–146, 2019.

[30] EU KLEMS, Growth and Productivity Accounts, Statistical Database, 2021.

[31] OECD, *OECD Digital Economy Outlook*. Paris, France: OECD Publishing, 2020.