My Advisor, Her AI and Me: Evidence from a Field Experiment on Human-AI Collaboration and Investment Decisions

Xitong LI

HEC Paris

lix@hec.fr

(joint work with C. Yang, K. Bauer, and O. Hinz)

Is human intervention needed in AI-based Systems?

- Bringing human knowledge in algorithm training could increase the performance of algorithms (Fügener et al., 2021; Rahwan et al., 2019; Raisch and Krakowski 2021)
- Algorithm aversion (e.g., Longoni et al. 2019)
- Limited accountability on an algorithm's output (Buckley et al., 2021)
- Regulatory requirements
    - GDPR Article 22
    - The AI Act of European Commission

THE WALL STREET JOURNAL

Home World U.S. Politics E

CIO JOURNAL Content by Deloitt

The Wall Street Journal news department was not involved in pr

SHARE

COGNITIVE | FUTURE OF WORK | TALENT

AI, R

Google

Fa

hu

Subscribe | Sign In

FINANCIAL TIMES

US COMPANIES TECH MARKETS CLIMATE OPINION WORK & CAREERS LIFE & ARTS HTSI

Data protection + Add to myFT

UK suggests removing human review of AI decisions in data protection laws

Excising Article 22 of Brussels regulations would alarm privacy campaigners

Article 22 of the GDPR

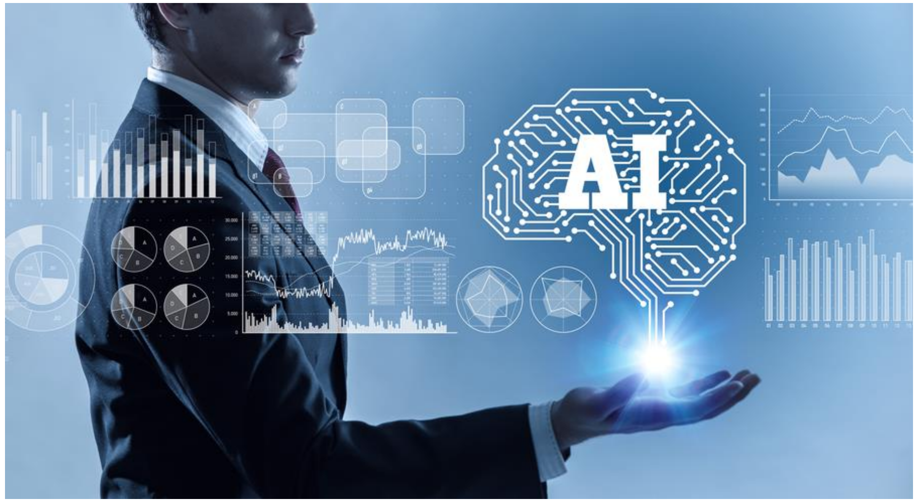
Article 22.1 of the GDPR sets an explicit prohibition to solely rely on automated processing for decisions which produce legal effects or similarly affects data subjects. **The GDPR introduces here human intervention as an essential component in the decision-making process**. May 31, 2022

https://trilateralresearch.com › research-highlights › huma...

Human intervention and human oversight in the GDPR and AI ...

Advice from Human-AI Collaborative Systems

Example 1: Medical Diagnosing
Physicians diagnose and advice patient with the support of AI systems, yet, having the final say



Example 2: Hybrid Financial Services
Bank advisors make recommendations to customers using their own experience and output from AI

Literature Gap

- Considerable literature on humans taking advice from other humans (experts/crowd)

- Growing literature on humans taking advice from AI

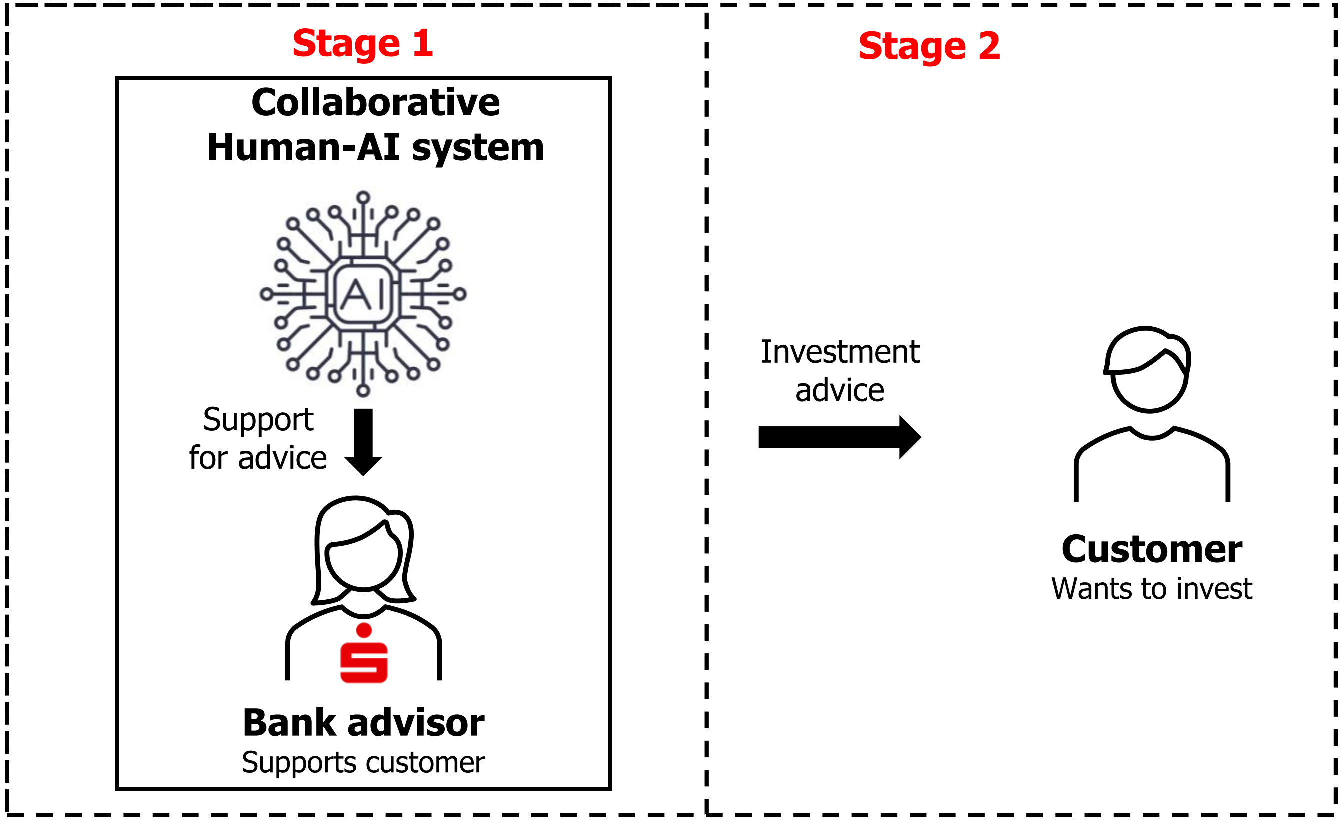
- **Little on humans taking advice from human-AI collaboration**
  - **especially, the value of human intervention in human-AI collaboration on both the production and consumption side** of AI-assisted services

Research Context & Research Questions

- Collaborated with one of the biggest savings banks in Germany
- Planned to offer personal loans as a new investment product

Production Side

Would allowing human experts to have the final say with the AI output improve or compromise advice quality?



Consumption Side

Would human-AI collaborative advice be more persuasive, compared to pure AI advice?

Production of Investment Advice (I)

➢ AI Advice

- We developed a deep learning algorithm to predict the default likelihood of real personal loans from LendingClub
- Using seven piece of information available
  - ✓ e.g., loan amount, APR, borrower’s income
- 90% as training data (over 1m loans) and 10% as test data
- Comparable prediction accuracy: 73%
- AI advice has two pieces:
  - ✓ Risk assessment: extremely low risky to extremely high risky
  - ✓ Investment recommendation (Yes/No)

Production of Investment Advice (II)

➢Human-AI Collaborative Advice

- A selected set of 24 personal loans (balanced in risk level and default)
- 27 expert bankers
- Each banker was presented with randomly selected 10 out of the 24 loans
- Each banker was asked to provide investment advice
  - ✓ one before and one after receiving the AI output
- Bankers' investment recommendations are incentivized (a payoff of 10€)
- The most frequent final investment recommendation and risk assessment are used as the human-AI collaborative advice
- A randomly selected 10 out of the 24 loans as investment opportunities for end customers

Production-Side Results

- ➢DV: Whether investment recommendation is correct
- ➢Independent var: whether it is produced after receiving AI output

Table A1: AI improves the quality of individual bankers' investment recommendations.

	Loans selected for customers (1)	All loans (2)
After receiving AI's output	0.647**	0.310*
	(0.312)	(0.160)
Observations	198	476
Banker fixed-effects	Yes	Yes
Loan fixed-effects	Yes	Yes
Order fixed-effects	Yes	Yes

Note: Standard errors clustered at the banker level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable is the investment recommendation accuracy. The reduced number of observations is due to perfect predictions of the dependent variable by certain loans or bankers after including the extensive sets of fixed effects.

Production-Side Results (Cont.)

- AI advice vs. human-AI collaborative advice
  - Compare the quality (predication accuracy) of the investment recommendations from the pure AI with that from the human-AI collaboration
  - For both types of recommendations, seven out of ten are correct

- **Key Results**
  - **(Good) AI could improve the quality (prediction accuracy) of financial service provided by bankers**
  - **Allowing bankers to have the final say with the AI output *does not* compromise advice quality**

Field Experiment on the Consumption Side

- Experimental Conditions
  - **AI-only condition** (benchmark): pure AI advice
  - **Human-AI collaborative advice**
    - ✓ Most frequent final investment recommendation across 27 bankers
  - **Human-only condition:** (perceived) human advice
    - ✓ Identical to the human-AI collaborative advice <u>without revealing bankers' use of AI</u>

Note: We find the advice quality of the three conditions is identical

Estimation Specification

$$V_{i,l} = \beta_1 AdviceInvest_{i,l} \times HumanAIAdvisor_i +$$
$$\beta_2 AdviceInvest_{i,l} \times HumanAdvisor_i +$$
$$\beta_3 AdviceInvest_{i,l} +$$
$$\alpha_1 HumanAIAdvisor_i + \alpha_2 HumanAdvisor_i + \gamma_1 X_{i,l} + \gamma_2 X_i + \phi_l + \eta_i^t$$

- DV: Final investment decision (0/1)
- AdviceInvest: Advise to invest or not (-1/1)
- Baseline: AI advice (AI-only condition)
- β_3 measures the alignment under AI-only condition
- β_1, β_2 measure the extent to which the alignment is higher or lower under human-AI and human-only conditions, respectively, compared to the AI-only condition

DV: Final Investment Decision
Logistic Regression

Human-in-the-loop leads to a higher degree of alignment with the advice!

	(1)	(2)
Human-AI × AdviceInvest	0.462**	0.462*
	(0.230)	(0.267)
Human-only × AdviceInvest	0.877***	0.877***
	(0.240)	(0.272)
AdviceInvest	0.815***	0.815***
	(0.237)	(0.283)
Human-AI	-0.246	-0.246
	(0.186)	(0.178)
Human-only	-0.364*	-0.364*
	(0.197)	(0.212)
AdvisedRiskAssess	-0.493***	-0.493***
	(0.157)	(0.174)
InitInvest	0.844***	0.844***
	(0.093)	(0.099)
InitRiskAssess	-0.104**	-0.104**
	(0.051)	(0.049)
GapInitAdvisedInvest	0.048	0.048
	(0.085)	(0.080)
GapInitAdvisedRiskAssess	-0.089	-0.089
	(0.058)	(0.059)
Age	-0.014***	-0.014***
	(0.004)	(0.004)
RiskPref	0.013	0.013
	(0.035)	(0.030)
Observations	1369	1369
Loan fixed-effects	Yes	Yes
Date fixed-effects	Yes	Yes
Std Err. clustered at ind-customer level	No	Yes

Positive effect of human-in-the-loop when customers face more risky investments

	More Risky Investments (1)	Less Risky Investments (2)
Human-AI × AdviceInvest	0.998*	0.399
	(0.553)	(0.326)
Human only × AdviceInvest	1.222**	0.917***
	(0.571)	(0.349)
AdviceInvest	1.082*	0.547
	(0.587)	(0.536)
Human-AI	-1.189***	0.183
	(0.371)	(0.249)
Human only	-2.053***	0.415
	(0.422)	(0.313)
Observations	557	792
Loan fixed-effects	Yes	Yes
Date fixed-effects	Yes	Yes
Ind-loan level ctrls	Yes	Yes
Ind-level ctrls	Yes	Yes

Standard errors clustered at the investor level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Positive impacts on downstream outcomes: payoffs

- Human-in-the-loop leads to higher payoffs for end customers

	Whole Sample	More Risky Investments	Less Risky Investments
	(1)	(2)	(3)
Human-AI	0.234***	0.535***	0.113
	(0.068)	(0.132)	(0.075)
Human only	0.208***	0.544***	0.044
	(0.070)	(0.144)	(0.081)
Observations	1369	567	802
Loan fixed-effects	Yes	Yes	Yes
Date fixed-effects	Yes	Yes	Yes
Ind-loan level ctrls	Yes	Yes	Yes
Ind-level ctrls	Yes	Yes	Yes

Standard errors clustered at the investor level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Investigation on Underlying Mechanisms

- Elaboration likelihood Model (ELM) suggests possible factors:
  - **Central route**: belief in the true advice quality
    - ✓ belief in advice quality
    - ✓ cognitive trust
  - **Peripheral route**: simple cues in the persuation context that do not change the advice content
    - ✓ advisor acccountablity
    - ✓ tolerance of wrong recommendations
    - ✓ social influence
    - ✓ emotional trust

Online Controlled Experiment

- Purpose
  - Replication on the findings of the field experiment
  - Underlying mechanisms

- Design
  - A fixed set of 8 loans
  - 300 German-speaking subjects on Prolific
  - Each was invited for 5 investment opportunities
  - Subjects' investment decisions are incentivized
  - Only AI advice used but with different advisor "labels"

Evidence for Underlying Mechanisms

	AdviceQuality	Accountability	SocialInfluence
	(1)	(2)	(3)
Human-AI	-0.055	-0.015	0.282*
	(0.178)	(0.100)	(0.157)
Human only	0.268	0.126	0.360**
	(0.190)	(0.103)	(0.165)
Observations	1500	1500	1500
Loan fixed-effects	Yes	Yes	Yes
Invest. order fixed-effects	Yes	Yes	Yes
Ind-loan level ctrls	Yes	Yes	Yes
Ind-level ctrls	Yes	Yes	Yes

Note: Standard errors clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Ind-level control include *Age*, *Female*, *Nationality*,

Key Findings

- Show the value of human intervention in AI-based service solutions
  - In our context, allowing humans to have the final say with the AI output
- Production-side value
  - Human intervention ***does not*** compromise service quality (prediction accuracy)
  - Higher level of accountability
- Consumption-side value
  - Higher level of persuasion (alignment with the advice)
  - Higher persuasive effectiveness of human-AI collaborative advice leads to higher end customers' welfare
- Driving factor: perceived social influence exerted by human bankers

Thanks for Your Attention!

Xitong LI
HEC Paris
lix@hec.fr