Plan Overview

A Data Management Plan created using DMPonline

Title: Eficacia de la colaboración entre humanos y algoritmos en la gestión de operaciones

Creator: Jeeva Somasundaram

Principal Investigator: Jeeva Somasundaram, Matthias Seifert

Contributor: Daniel Corsten, Elena Revilla, Lea Reusch

Affiliation: IE University

Funder: Science Europe

Template: Science Europe Template

ORCID iD: 0000-0001-5510-8447

Project abstract:

Algorithms increasingly support or automate operational decisions, yet their value depends on how people use, adjust or override them. HACOM studies when and how human–algorithm collaboration improves operational performance. The project combines controlled online and laboratory experiments, field experiments with industry partners, and analysis of proprietary operational data. It pursues four objectives: (1) how decision makers respond to risk-averse algorithmic recommendations in inventory (newsvendor) decisions; (2) whether the source of advice, human or algorithm, changes how strongly and persistently people rely on it; (3) how AI-augmented versus automated demand forecasting affects forecast accuracy and inventory decisions in a global consumer-goods supply chain; and (4) how algorithmic recommendations and human behaviour interact in online grocery retailing, including product substitutions, order picking and gig-worker task allocation. Allied studies examine digital behaviour, AI-assisted idea generation, and algorithm-enabled transparency in humanitarian and textile supply chains. Data generated include anonymised experimental data, confidential company and platform data, and qualitative interview data.

ID: 216456

Start date: 01-09-2023

End date: 31-08-2026

Last modified: 25-11-2026

Grant number / URL: PID2022-138667OB-I00

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Eficacia de la colaboración entre humanos y algoritmos en la gestión de operaciones

Data description and collection or re-use of existing data

How will new data be collected or produced and/or how will existing data be re-used?

The project generated data through four routes. (1) Online behavioural experiments and surveys run on Prolific and Amazon Mechanical Turk (Objectives 1–2 and allied studies), including incentivised inventory-decision (newsvendor) tasks with algorithmic advice, studies on smartphone use, and experiments on AI-assisted idea generation run through a purpose-built web application. (2) A field experiment with a global consumer-goods company and one of its online-retail customers, using the company's operational forecasting and inventory records (Objective 3). (3) Re-use of existing proprietary operational data supplied by an anonymous European online grocery platform under a non-disclosure agreement (Objective 4). (4) Qualitative observations and recorded interviews in textile factories in Bangladesh (allied stream). No publicly available datasets were re-used.

What data (for example the kinds, formats, and volumes) will be collected or produced?

- Experimental and survey data: anonymised participant-level decisions and questionnaire responses (CSV/Excel, Stata .dta, R files), including 400 Amazon Mechanical Turk participants in Study 1a and multiple Prolific samples across the follow-up and allied studies.
- Company field data (Objective 3): forecasting, judgmental-adjustment and inventory records (Excel/Stata).
- Platform data (Objective 4): a customer and substitution dataset (241,505 orders; 1.51 million item observations; 57,360 customers) and a picking/worker-performance dataset (more than 2.5 million orders over 33 months), plus a randomised field experiment (47,141 customers, 185 stores). Stored as CSV/database extracts.
- Qualitative data: audio recordings of interviews, verbatim transcripts and field notes (MP3/M4A, Word/PDF).
- Code: analysis scripts (Stata, R, Python) and the source code of the experimental application.

Total volume is moderate (below 50 GB).

Documentation and data quality

What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?

Each dataset is accompanied by a README describing the study, the data-collection method and period, the sample and the inclusion/exclusion criteria; a codebook defining every variable, its units and its coding; the experimental instructions and survey instruments; and the analysis scripts that reproduce the reported results. Files follow a consistent folder structure (raw / cleaned / analysis / output) and naming convention by study.

What data quality control measures will be used?

Online studies used attention checks, comprehension questions, duplicate-participant screening (platform IDs and IP checks) and pre-specified exclusion rules. Experimental software was pilot-tested before launch. Company and platform data were validated against summary statistics supplied by the partners, and anomalies were resolved with them. Cleaning steps are scripted rather than done by hand, so every transformation from raw to analysed data is reproducible. Co-authors cross-checked key analyses, and journal peer review provided further scrutiny.

Storage and backup during the research process

How will data and metadata be stored and backed up during the research process?

Data are stored on IE University's institutional cloud storage (Microsoft OneDrive/SharePoint), which is automatically backed up and version-controlled. Confidential partner data are stored only in restricted-access folders and are not copied to personal devices beyond the encrypted, password-protected project laptops. Code is version-controlled. The experimental application's database

(Heroku) was exported and archived on institutional storage at the end of data collection. Shareable final datasets are also deposited on the Open Science Framework (OSF).

How will data security and protection of sensitive data be taken care of during the research?

Access is restricted to the researchers working on each study and is managed through institutional accounts with multi-factor authentication. Personal identifiers (e.g. Prolific/MTurk IDs) are separated from response data and deleted or pseudonymised after payment and quality checks. Platform and company data were received already anonymised or pseudonymised, contain no directly identifying customer information, and are handled according to the partners' non-disclosure agreements. Interview audio recordings and transcripts are stored only on restricted-access institutional storage; transcripts are pseudonymised, and the key linking codes to identities is stored separately. Recordings are deleted once pseudonymised transcripts are verified, or at the latest at the end of the retention period.

Legal and ethical requirements, codes of conduct

If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

Personal data were processed in line with the EU General Data Protection Regulation (GDPR) and Spain's Organic Law 3/2018 on Data Protection (LOPDGDD). Participants received information sheets and gave informed consent before taking part, including specific consent for audio recording of interviews. Data minimisation was applied: only data needed for the research were collected. Identifiers were removed as early as possible, and only anonymised data are used in analyses and publications. IE University's Data Protection Officer is the point of contact for data-protection matters.

How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

Experimental and survey data generated by the team are owned by IE University and the partner universities of the researchers involved. Company and platform data remain the property of the providing organisations and are used under non-disclosure/data-use agreements (including a 2023 NDA with the grocery platform), which restrict sharing and require anonymity of the partner. Applicable legislation: GDPR, LOPDGDD, Spanish intellectual property law (Real Decreto Legislativo 1/1996), and the open-science obligations of Spain's Science, Technology and Innovation Law (Ley 14/2011, as amended by Ley 17/2022).

How will possible ethical issues be taken into account, and codes of conduct followed?

All studies involving human participants received ethical approval through IE University's research-ethics procedures before data collection. Participants gave informed consent, were paid fairly on the online platforms, and could withdraw at any time. No study used deception. Interview participants in Bangladesh consented to audio recording, and the anonymity of workers and firms was protected throughout. The team follows the European Code of Conduct for Research Integrity (ALLEA).

Data sharing and long-term preservation

How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

Anonymised experimental and survey data, codebooks and analysis code are shared openly on OSF when the corresponding article is published (or on acceptance). Data from the Objective 3 field experiment and the Objective 4 platform data cannot be shared publicly because of confidentiality agreements with the partner companies; for these, the code, variable descriptions and aggregated results are shared, and access to the underlying data may be possible on request, subject to the partners' approval. Interview recordings and full transcripts are not shared, to protect participants' anonymity; anonymised excerpts appear in publications.

How will data for preservation be selected, and where will data be preserved long-term (for example a data

repository or archive)?

All datasets underlying published results are preserved: the final cleaned data, codebooks and replication code. Shareable data are deposited on the Open Science Framework (OSF), and the articles' data-availability statements link to them. Confidential partner data and interview materials are retained on secure institutional storage for at least five years after the end of the project (or as the partner agreements require) and then deleted.

What methods or software tools will be needed to access and use the data?

Data are provided in open, non-proprietary formats (CSV, plain text and PDF) alongside native files. Analysis scripts run in Stata, R or Python. R and Python are free, and Stata scripts are documented so that they can be reproduced in other software. No specialised tools are required.

How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

Each dataset deposited on OSF receives a DOI. The DOI is cited in the corresponding article's data-availability statement and listed in the project's final report.

Data management responsibilities and resources

Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

The principal investigators, Prof. Matthias Seifert (IP1) and Prof. Jeeva Somasundaram (IP2) of IE University, are responsible for data management overall. Each study lead manages the data for their study: Elena Revilla (Objective 3), Daniel Corsten (Objective 4) and Lea Ruesch (allied humanitarian and supply-chain studies). IE University's library/research office supports data deposit and the Data Protection Officer supports GDPR compliance.

What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Data management is part of the researchers' working time and was planned as part of each study (documentation, cleaning scripts, preparation for deposit). The project paid for secure data-collection platforms and research software (Prolific, Amazon Mechanical Turk, Stata, the experiment application's hosting). Deposit on OSF is free, so it needs no extra budget. Using open formats, codebooks, persistent identifiers and clear licences (CC BY 4.0 for shareable data) makes the data findable, accessible, interoperable and reusable.

Planned Research Outputs

Dataset - "Algorithmic advice experiments – newsvendor studies (HACOM-ALGO)"

Anonymised participant-level data from online experiments on Amazon Mechanical Turk and Prolific in which participants made inventory (newsvendor) ordering decisions with or without algorithmic advice of varying risk aversion, and with advice framed as coming from a human or an algorithm (Objectives 1–2). Includes order decisions, profits, advice uptake, treatment assignment and survey measures. CSV and Stata files with README, codebook and experimental instructions.

Dataset - "Digital behaviour and AI ideation experiments"

Dataset - "Humanitarian social-media message experiments"

Anonymised data from online experiments testing how the content of social-media messages posted by humanitarian workers (authenticity, emotional tone) affects information sharing and donation behaviour. Includes participant choices, treatment assignment and survey measures. CSV files with README, codebook and stimuli.

Dataset - "AI-augmented demand-planning field data"

Forecasting, judgmental-adjustment and inventory records from a field experiment with a global consumer-goods company and one of its online-retail customers, comparing AI-automated and AI-augmented (human-adjusted) forecasting (Objective 3). Contains no personal data. Commercially confidential under the company agreement and not shared; aggregated results and analysis code are made available.

Dataset - "Online grocery platform data"

Dataset - "Bangladesh factory interviews"

Audio recordings, verbatim transcripts and field notes from observations and interviews in textile factories in Bangladesh on how sustainability data are captured and shared for transparency platforms and digital product passports. Contains personal data. Not shared, to protect participants' anonymity; anonymised excerpts appear in publications.

Dataset - "Analysis code and experiment application"

Stata, R and Python scripts that clean the data and reproduce all analyses reported in the project's publications, together with the source code of the web application used for the AI-ideation experiments. Shared openly on OSF alongside the corresponding datasets.

Planned research output details

Title	DOI	Type	Release date	Access level	Repository(ies)	File size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Algorithmic advice experiments – newsvendor studie ...		Dataset	Unspecified	Open	None specified		None specified	None specified	No	No
Digital behaviour and AI ideation experiments		Dataset	Unspecified	Open	None specified		None specified	None specified	No	No
Humanitarian social-media message experiments		Dataset	Unspecified	Open	None specified		None specified	None specified	No	No
AI-augmented demand-planning field data		Dataset	Unspecified	Open	None specified		None specified	None specified	No	No
Online grocery platform data		Dataset	Unspecified	Open	None specified		None specified	None specified	No	No
Bangladesh factory interviews		Dataset	Unspecified	Open	None specified		None specified	None specified	No	No
Analysis code and experiment application		Dataset	Unspecified	Open	None specified		None specified	None specified	No	No