OPEN LOOP

Artificial Intelligence Act:

A Policy Prototyping Experiment

Towards Informed AI Interactions:
Assessing the Impact of Notification Styles on User Awareness and Trust

52

NORBERTO DE ANDRADE
LAURA GALINDO
ANTONELLA ZARRA
JESSICA HEAL
SARAH ROM

We would also like to thank the Open Loop's partners and observers from business, academia, civil society, international institutions and governmental authorities:

Members of European Parliament

- **Eva Maydell**
  ITRE Rapporteur of the AI Act (BG, EPP)

- **Ivan Štefanec**
  President of SME Europe (SK, EPP)

International Institutions / Governmental Authorities

- **Alessandro Fusacchia**
  Former Member of the Italian Parliament, Coordinator of Parliamentary Intergroup on AI

- **Andras Hlacs**
  AI Policy Analyst, OECD.AI, Organisation for Economic Co-operation and Development (OECD)

- **Işıl Selen Denemeç**
  LL.M Head of Legal Department, Digital Transformation Office of the Presidency of the Republic of Türkiye

- **Karine Perset**
  Head of AI Unit and OECD.AI, Organisation for Economic Co-operation and Development (OECD)

- **Luca Carabetta**
  Former Member of the Italian Parliament

- **Richard Nevinson**
  Head of Digital Economy, Information Commissioner's Office (ICO)

- **Zümrüt Müftüoğlu**
  Expert, Digital Transformation Office of the Presidency of the Republic of Türkiye

Academia / Think Tanks

- **Andrea Bertolini**
  EURA Centre, Scuola Superiore Sant' Anna Pisa

- **David Osimo**
  Lisbon Council

- **Eduard Fosch**
  Leiden University

- **Evert Stamhuis**
  DIGOV Centre Rotterdam University

- **Fabiana di Porto**
  University of Salento and LUISS University

- **Giovanni Sartor**
  University of Bologna and European University Institute

- **Johann Laux**
  Oxford Internet Institute

- **Joshua Ellul**
  University of Malta and Former chairman, Malta Digital Innovation Authority

- **Klaus Heine**
  DIGOV Centre Rotterdam University

- **Nicolaos Voros**
  University of the Peloponnese

- **Risto Uuk**
  The Future of Life Institute

- **Virginia Dignum**
  Umeå University

OPEN LOOP

Industry Associations

Asociación Española de Startups

CRCA

DEVELOPERS ALLIANCE

infobalt LIETUVA

InnovUp ITALIAN INNOVATION & STARTUP ECOSYSTEM

sme connect

Start up Lisboa

Civil Society

DATASPHERE INITIATIVE

THE FUTURE SOCIETY

Pop Ai

PRIVACY NETWORK

OPEN LOOP

Artificial Intelligence Act: A Policy Prototyping Experiment

Overview of the Open Loop Program on the AI Act and the stakeholders involved

Operationalizing the Requirements for AI Systems

Regulatory Sandboxes

Taxonomy of AI Actors

1 — *Part I* — **The Open Loop Forum**

2 — *Part II* — **The Deep Dives**

3 — *Part III* — **Transparency Obligations for AI-Human Interaction**

You are here

Legend

- AI Companies
- Business Associations
- End Users
- Policymakers
- Civil Society
- Academia

Contents

Executive summary

Executive summary

As part of the Open Loop program on the Artificial Intelligence Act, we tested one of the requirements of Article 52(a) (on transparency obligations for AI systems interacting with individuals) of the proposed regulation to assess when and how individuals should be informed when they are interacting with an AI system.

We conducted an online survey with a sample of 469 participants from five European countries (Spain, France, Germany, United Kingdom and Sweden). Via this online survey, participants were exposed to videos of two different AI-powered systems, a chatbot and a news app. Different styles of AI notifications were presented to the participants: i) no notification, ii) content-integrated notification and iii) notification banner.

In this study, participants' reactions to different notification styles and depths of information were measured based on the following variables:

i) Perception and Understanding of Interacting with AI: Participants' perception and understanding of their interactions with AI systems were assessed. This variable aimed to determine how well individuals recognized and comprehended their engagement with AI.

ii) Trust and Sense of Agency: The study examined participants' level of trust and their sense of agency over the AI systems. This variable explored the extent to which individuals felt in control and confident in their interactions with AI.

The findings revealed the following:

a. Perceived and Understanding of Notification:

- Participants who saw the Notification Banner style had a higher rate of perceiving and understanding it (70%) compared to those who saw the Content-integration style (51%).

- Participants who were already aware of other AI-powered products were more likely to notice and understand the notification.

- Participants' digital literacy and knowledge of advanced technologies did not significantly affect their awareness of interacting with AI.

b. Understanding of Interacting with AI:

Participants who perceived and understood the notification banner had a strong understanding of interacting with AI in the presented application (80.8%).

Participants who perceived and understood content-integrated notifications also had a good understanding that they were interacting with an AI system (76.5%).

Even without perceiving the notification, a majority of participants (57.1%) understood that they were interacting with AI.

Some participants who were familiar with using AI did not expect to interact with AI in the applications presented (17.3%).

c. Trust and Sense of Agency:

Participants with higher awareness of interacting with AI in AI-powered products showed greater trust in the applications presented.

Participants who expected to interact with AI in similar products also exhibited higher trust in the application.

Participants' awareness and understanding of the notification did not significantly affect their sense of agency and trust in the applications.

The type of application and the depth of information presented did not significantly impact participants' sense of agency and trust.
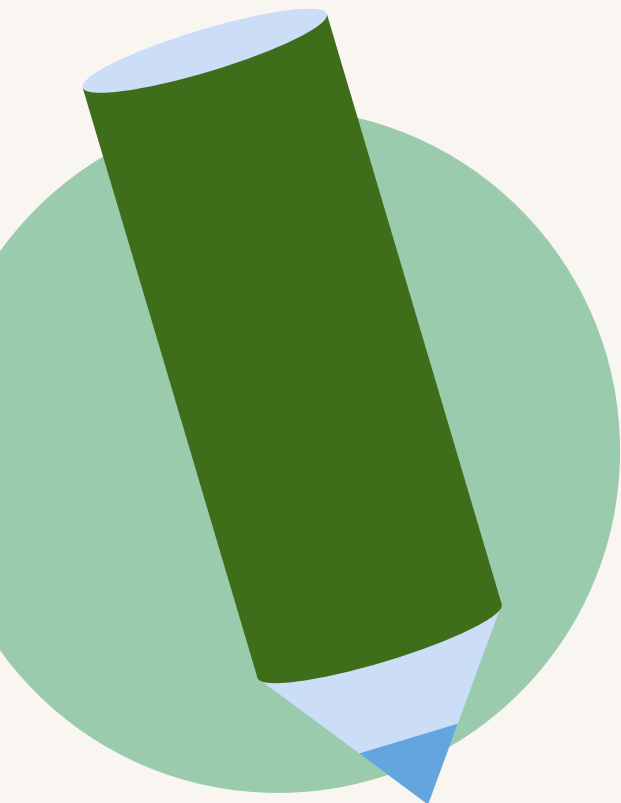
1

Introduction

Article 52 of the Artificial Intelligence Act (AIA) establishes certain transparency obligations for AI providers and AI users whose AI products/services interact with natural persons.[1] Article 52(1) AIA reads:

> *"Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate and prosecute criminal offences, unless those systems are available for the public to report a criminal offence."*

The explanatory memorandum to the AIA states that the goal of Article 52 is to inform people of the fact that they are interacting with an AI, in particular to avoid risks of manipulation and allow individuals to make informed choices.[2]

To test Article 52(1) AIA, as part of the Open Loop Program on the EU AI Act,[3] we conducted a survey in five countries in Europe (Germany, France, Italy, United Kingdom, Sweden) to test the effects of different types of notifications on individuals' awareness, understanding, trust, and sense of agency over AI.
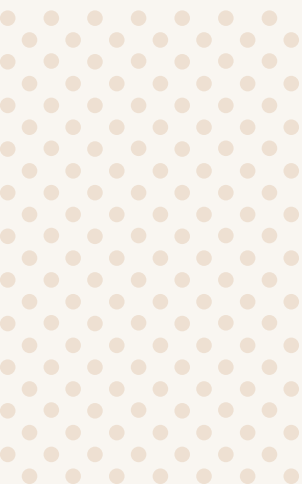
2

Background
to the research
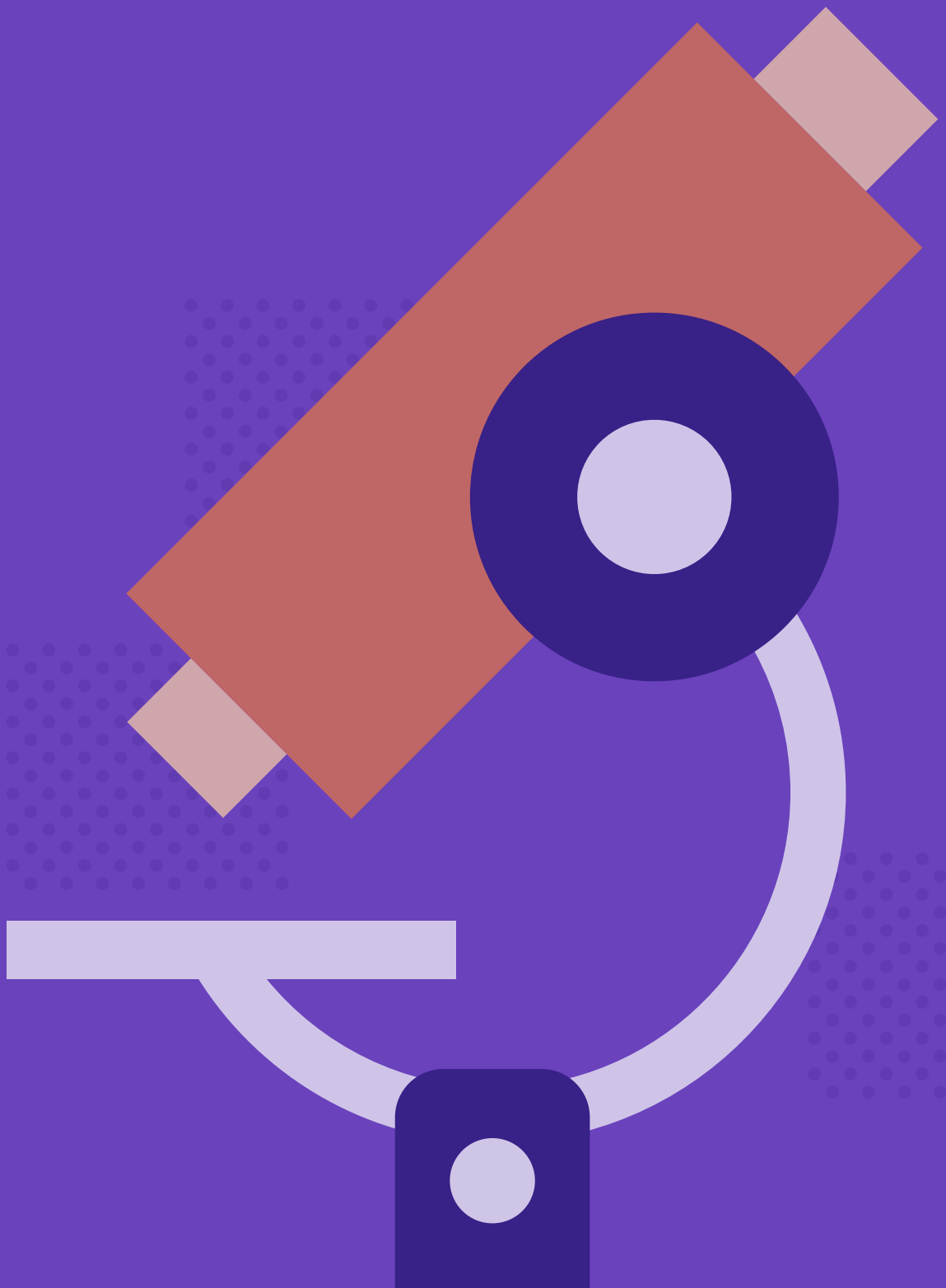
Article 52 in the AI Act: theory of change

Developing a theory of change enables the mapping of the requirements that help achieve a certain policy goal. The figure below illustrates the theory of change for Article 52:

AIA Requirement 1

AI providers design their AI system in such a way that it is clear for individuals that they interact with an AI system (52^1 AIA)

AIA Requirement 2

Users inform individuals when a biometric /emotion recognition AI system is used (52^2 AIA)

AIA Requirement 3

Users label artificial content as such (52^3 AIA)

Outcome 1

Individuals are aware they are interacting with an AI

Outcome 2

Individuals are aware that an AI system using biometric /emotion recognition is used

Outcome 3

Individuals can recognize artificially generated content

Policy Goal

1 Avoid risks of manipulation by/through AI
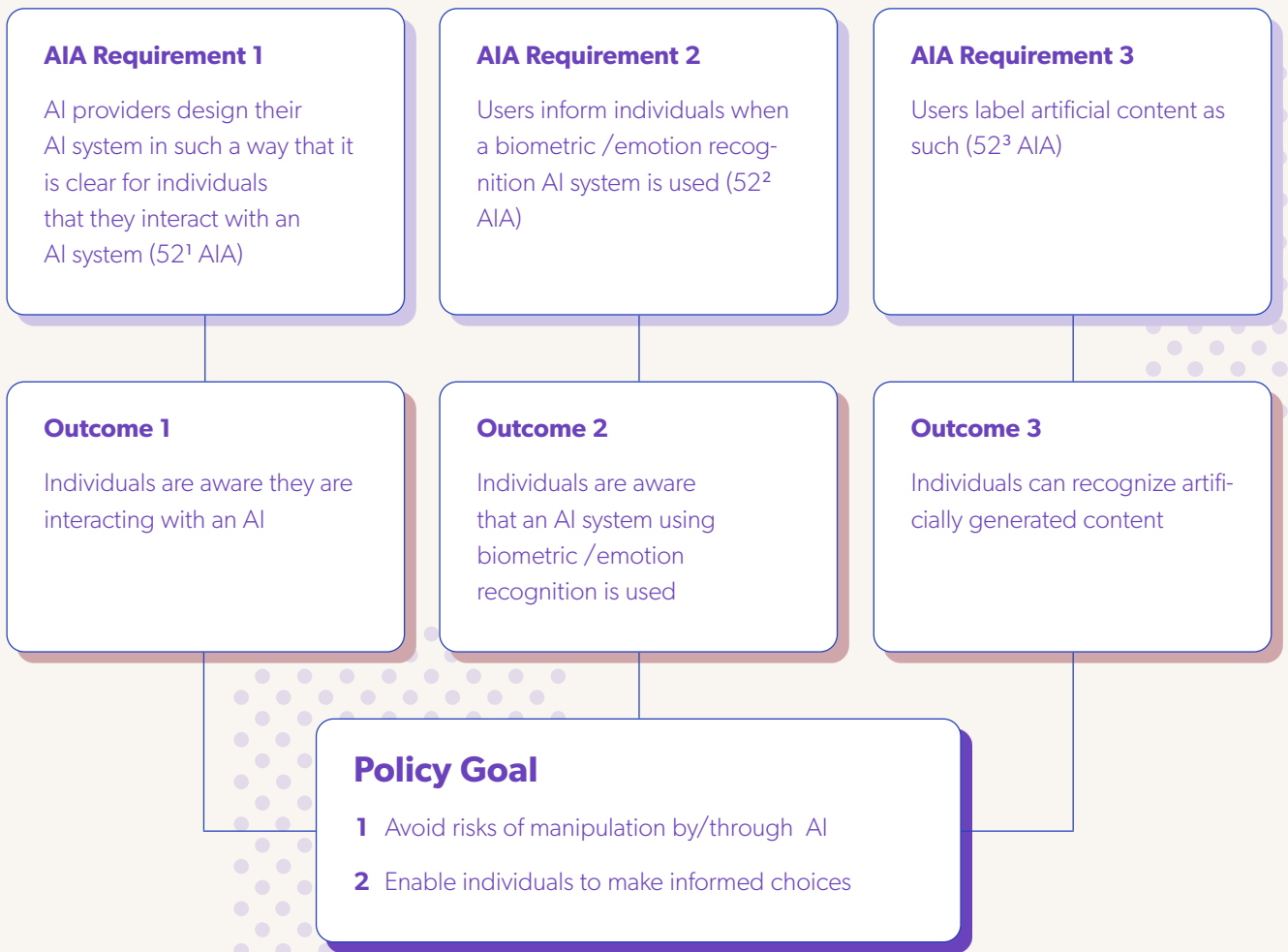
2 Enable individuals to make informed choices

Figure 1 Theory of Change - Art 52 on transparency obligations for AI providers and users in the AIA

In this theory of change we show how the requirements set forth in Article 52 (paragraphs 1–3) should lead to the desired outcomes, which contribute to the policy goal.

The policy goals underlying Article 52 are twofold. Firstly, it aims to prevent manipulation by or through AI systems, safeguarding individuals from impersonation, deception, and other ill-intentioned actions. Secondly, it seeks to empower individuals to exercise control, opt-out, or make choices based on their awareness of interacting directly or indirectly with AI.

These policy goals may mitigate the potential risks a person is exposed if unaware that they are interacting with AI. This potential risk can be as a result of bad intent (e.g. impersonation or deception) but can also simply be the result of a person not recognizing that they are dealing (directly or indirectly) with an AI.

To avoid these issues, Article 52 establishes transparency obligations for AI providers and AI users.

13

Transparency Obligations

1 **Design Clarity:** AI providers are required to design their systems in a manner that makes it evident to individuals that they are engaging with an AI system (Article 52(1)).

2 **Emotion Recognition and Biometric Systems:** AI users must inform individuals when they are exposed to emotion recognition systems or biometric categorization systems (Article 52(2)).

3 **Labeling Artificial Content:** AI users are obligated to label artificial content, such as deep fakes, as artificial (Article 52(3)).

In the context of this part of the Open Loop policy prototyping program on the EU AI Act, we have focused on Article 52(1) AIA, the requirement that AI systems should be designed in such a way that it is clear to individuals that they are interacting with AI. In practice, this requirement will most likely entail that some kind of notice is given to individuals. We have focused on this notification.

Purpose of the research

The main purpose of testing Article 52 AIA is to assess when and how individuals should be informed when they are interacting with an AI system. While Article 52 AIA sets forth that individuals should be informed, it does not make it explicit. According to the regulation, natural persons should be informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use. Our assumption is that it is not always obvious for all individuals that they are interacting with an AI system. That is why we wanted to test when it is clear for natural persons that they are interacting with an AI system. Not only did we want to test whether it was clear to an individual from the context alone that they were interacting with an AI system, we also wanted to test to what extent informing them via a notice had any impact on their level of awareness.

In particular, we explored the two following research questions (RQs):

> **RQ1: Does a notification increase awareness of individuals that they are interacting with an AI?[4]**

The hypothesis is that through a notification, individuals become aware that they are interacting with an AI. This awareness in turn will allow individuals to recognize and/or avoid manipulation and, in turn, to make informed choices.

> **RQ2: Does a notification lead to an increase in individuals' trust and agency over AI?[5]**

The legislator argues that once a user is aware of interacting with AI they will be able to make an informed choice to continue interacting with the AI. As such, the hypothesis is that notification will increase individuals' trust and agency.

Data and methodology

This exploratory research investigates in test conditions how individuals respond to different types of AI notifications. It provides preliminary insight into what could be the most effective methods for satisfying Article 52 AIA transparency obligations and lays the foundation for further research of the AIA in regard to information design.

Research Design

An online survey was conducted with a sample of 469 participants from five European countries.[6] Via this online survey, participants were exposed to videos of two different AI systems (a customer service chatbot called 'Chat App' and a newsfeed algorithm called 'News App').

Across these two AI systems, participants were either:

- i) Not notified of the application being powered by AI,

- ii) Notified through a notification banner, or

- iii) Notified through a content-integrated notification

For each application, three different videos were created (see following pages and Annex/Proto-types/Videos) – each containing a different style of notification or no explicit notification.
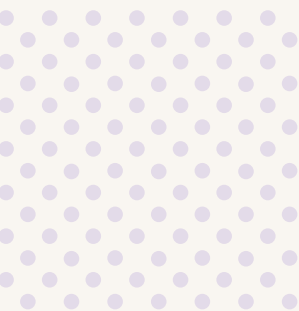
In **no notification versions**, only subtle hints allow the viewer to conclude that the application might require AI - for example "My Feed" suggests that the content is personalized. This sets the baseline to compare the effects of different notifications.

Both the **content-integrated** and the **notification banner** versions use similar wording and contain the phrase "AI-powered". One of these three options (notification styles) was randomly selected for the participants. For both applications, all notification versions were designed similarly in layout, font size, and shape to enable a comparison.

To help participants differentiate the two applications, two different colors were used. To ensure individuals refer to the correct user interface element when being asked for the notification, the notification elements were colored blue (Chatbot) and purple (Newsfeed).

In addition, for each of the three types of notification (i) no notification, ii) notification banner, and iii) content-integrated notification) a mock-up image of a helpdesk page providing *more information* on how AI is used in the specific application was designed.

This enables the explorative study of effects caused by three *depths of information:* i) No explicit information on the usage of AI, ii) only communicating that the provided functionality is AI-powered versus iii) briefly explaining the usage of AI to allow application-specific functionality.

Types of notifications

Version

No notification. (style i).

Participants were not explicitly notified of their interaction with an AI system. Only subtle hints could allow the participant to conclude this interaction.
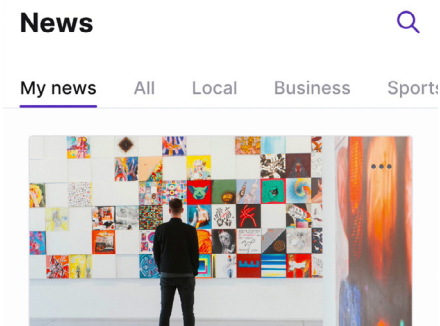
Chatbot

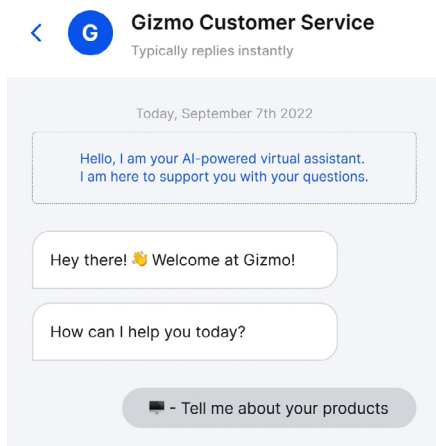Only Hints: "typically replies instantly", generic wording, providing answer options.

Gizmo Customer Service
Typically replies instantly

Today, September 7th 2022

Hey there! 👋 Welcome at Gizmo!

How can I help you today?

💻 - Tell me about your products

Newsfeed

Only Hints: News categories "My news", "local".

News

My news | All | Local | Business | Sport

Content-integrated. (style ii).

Participants were notified of their interaction with an AI system through communication within the content (news for News Feed and messages for Chatbot). The notification appears as part of the content.

Blue introduction message located in the message area, scrolling with messages.

Gizmo Customer Service
Typically replies instantly

Today, September 7th 2022

Hello, I am your AI-powered virtual assistant. I am here to support you with your questions.

Hey there! 👋 Welcome at Gizmo!

How can I help you today?

💻 - Tell me about your products

Purple message explaining that the personal content selection is AI-powered, scrolling with content.
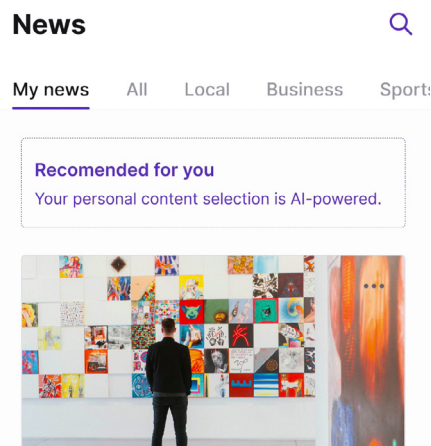
News

My news | All | Local | Business | Sport

Recomended for you
Your personal content selection is AI-powered.

Types of notifications

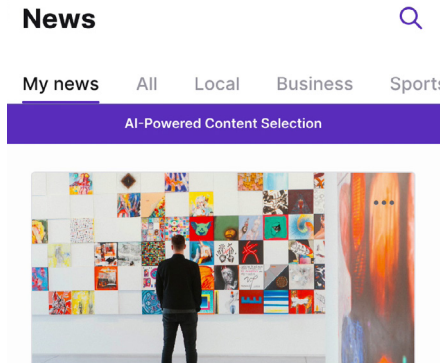| Version	Chatbot	Newsfeed
Notification banner. (style iii).	Blue notification bar located on top of the message area, not scrolling with messages.	Purple notification bar explaining that the personal content selection is AI-powered, not scrolling with content.
Participants were notified of their interaction with an AI system via a notification banner. This notification banner appeared as part of their user interface and had no proximity to the content of the AI system.		
More information.	Information on how the chatbot uses AI to conduct conversations.	Information on the criteria the AI content selection is based on.

Table 1. Types of notifications

Types of notifications
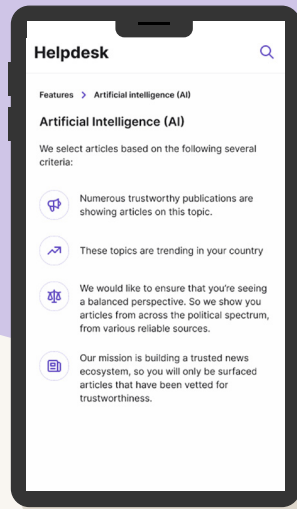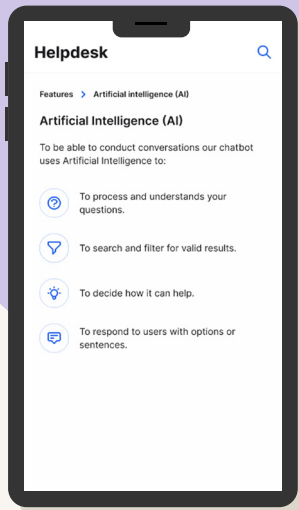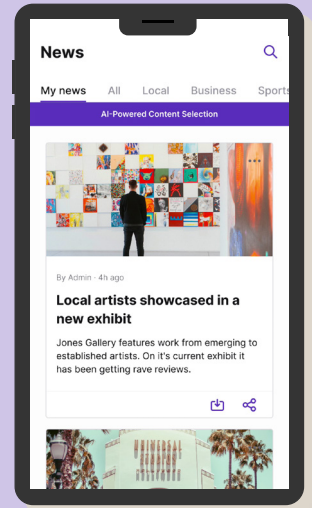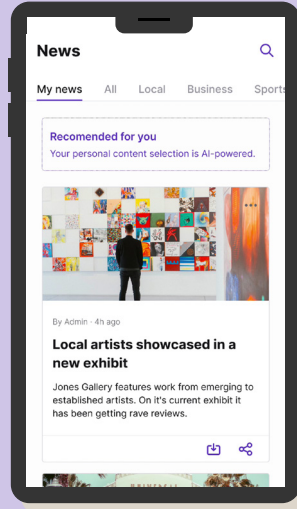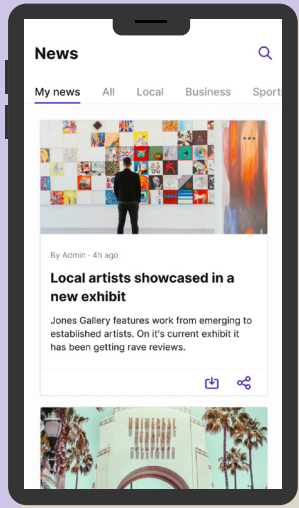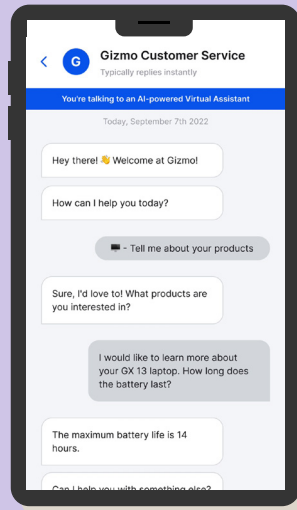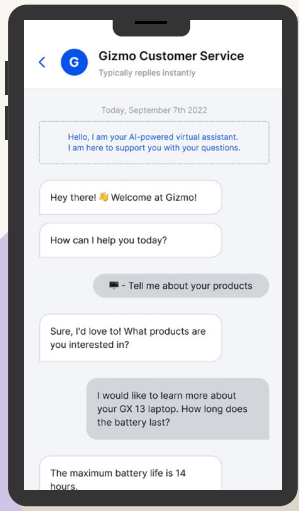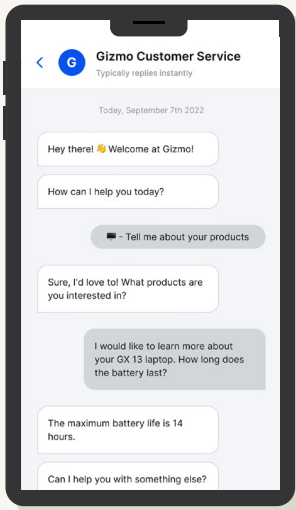
Variables

In this study, participants' reactions to different notification styles and depths of information were measured based on the following variables:

1 **Perception and Understanding of Interacting with AI:** Participants' perception and understanding of their interactions with AI systems were assessed. This variable aimed to determine how well individuals recognized and comprehended their engagement with AI.

2 **Trust and Sense of Agency:** The study examined participants' level of trust and their sense of agency over the AI systems. This variable explored the extent to which individuals felt in control and confident in their interactions with AI.

Participants' Characteristics:

To gain insights into the influence of participants' characteristics on their reaction to notifications, the following specific characteristics were considered:

1 **Digital Literacy:** Participants' level of digital literacy was taken into account. This variable aimed to assess the participants' familiarity and competence in using digital technologies.

2 **Awareness of Interacting with AI Systems:** Participants' pre-existing awareness of interacting with AI systems outside of the research context was examined. This variable aimed to understand if individuals would be able to recognize AI interactions in neutral settings.

Limitations of the research

While this study offers valuable insights, it is essential to acknowledge certain limitations that may impact the generalizability of the findings. These limitations include:

- Limited Sample and Survey Design: The study employed a sample size that was restricted in scope, and participants were presented with designs through a survey rather than experiencing real-world product interactions. Consequently, caution should be exercised when interpreting the results, and they should be viewed as preliminary indications rather than definitive conclusions.

- Focus on Article 52(1) AIA: The study concentrated specifically on AI notifications as a means to fulfill the transparency obligation outlined in Article 52(1) of the Artificial Intelligence Act (AIA). It did not encompass an examination of the other transparency obligations stated in Article 52, such as the notification of biometric emotion recognition systems (Article 52(2) AIA) or the mandatory disclosure of artificially generated content like deepfakes (Article 52(3) AIA).

3

Results & Highlights

Did participants perceive and understand the notification?

Results

- Participants who were shown the Notification Banner style (style iii) perceived and understood it in 70% of cases, while participants who were shown the Content-integration style (style ii) only perceived and understood it in 51% of cases.

- Participants who had pre-existing awareness of other AI-powered products (besides Chatbots and Newsfeeds) were more likely to both notice and understand the notification.

- Digital literacy (e.g. regular use of email) and knowledge of advanced technologies (e.g. block-chain technology) had no significant effect on the awareness of interacting with AI.

Discussion

- While the notification stood out from the design for both notification styles, it was still not perceived by a significant portion of the participants (30% and 49% respectively). This could be explained by the concept of 'banner blindness', a phenomenon in which website users consciously or unconsciously ignore banner-like information, such as ads, calls to action, or other promotional content, due to the perception that such content is irrelevant or intrusive.

- Despite this, the banner variant was perceived as more noticeable by participants. This may be due to the differences in design.

- From a design perspective, the notification banner style seems to yield the best results in terms of grabbing people's attention. This can be attributed to its more obtrusive visual design, in particular the contrasting background color.

- What is interesting to note is that participants that had prior experience / knowledge of interacting with AI were more likely to perceive and understand the shown notification. An explanation for this could be that participants with prior knowledge of AI interaction can better contextualize the information presented and are therefore more likely to consciously notice the notification. More general digital literacy had no significant effect, so it seems that the specific experience of interacting with AI is relevant in recognizing future interactions with AI.

Did participants understand that they were interacting with an AI system?

Results

- In 80.8% of the cases, participants who perceived and understood a notification banner showed a high understanding of interacting with AI in the presented application.

- In 76.5% of the cases, participants who perceived and understood the content-integrated notifications had a high understanding that they were interacting with an AI system.

- Even without perceiving the notification, 57.1% of participants understood that they were interacting with an AI system.

- A large group of participants seems to notice a difference in meaning between the use of and the interaction with AI: Approximately 17.3% of participants with a strong understanding of using AI did not assume they would be interacting with AI in the applications presented.

Discussion

- Participants that perceived and understood the notification (either style) had a greater understanding that they were interacting with an AI system, than those that didn't. In that sense, the notification seems to contribute to its intended goal, as it leads to a higher level of awareness/ understanding of interacting with AI. In particular for those that have previous experience with interacting with AI systems the notification leads to a higher understanding. This is likely due to the fact that they can better contextualize what the notification means.

- Still, even without perceiving or understanding the notification, a slight majority of the participants understood that they were dealing with an AI system. So, clearly there are other cues that lead people to deduce that they are interacting with an AI system.

- In the study, the notification let participants know that the application was "AI-powered". Interestingly, while participants inferred that this would mean the application was "using" AI, they did not expect to be "interacting" with AI.

Do the notifications provide individuals with trust and a sense of agency when interacting with an AI system?

Results

- Participants with higher awareness of interacting with AI when using AI-powered products show higher trust in the applications presented.

- Participants who expect to interact with AI in similar products show a greater trust in the application presented.

- The participants' awareness of interacting with AI when using AI-powered products shows no significant effects on their sense of agency.

- The participants' expectation to interact with AI in products similar to the displayed applications shows no significant effects on their sense of agency.

- Whether participants perceived and understood a notification shows no significant effects on their sense of agency and trust in the applications presented.

- The type of application presented to the participants shows no significant effects on their sense of agency and trust in the application presented.

- The depth of Information that was presented to the participants shows no significant effects on their Trust and their sense of Agency in the applications presented.

Discussion

- Participants have a higher trust in applications if they are aware of interacting with AI when using other AI-powered applications. However, this has no influence on their sense of agency.

- The level of trust or agency was not impacted by the type of application.

- Whether participants perceive and understand a notification, it does not seem to significantly affect their trust in the application, nor their sense of agency. Also, the depth of information provided had no significant influence. We conclude that the notification styles tested as well as the depth of the information provided have little impact on participants' sense of agency and trust in AI systems.

- While we assessed only whether being exposed to a notice increased individuals' agency and not the reasons behind, the lack of agency could be explained by the fact that the participants could not influence the interaction with AI (e.g. opting out to AI powered interactions).

Conclusions

The study findings indicate that the notification styles tested, as well as the depth of information provided, have minimal impact on participants' sense of agency and trust in AI systems. Prior experience with AI interaction and awareness of AI usage in other applications contribute to a higher understanding of interacting with AI. However, factors such as the type of application and the perception and understanding of the notification do not significantly influence participants' trust or sense of agency. The study suggests the importance of considering user agency and the need for further research to explore reasons behind the lack of agency in AI interactions.

4

Policy implications and recommendations

While more research is needed into the information requirements of Article 52(1) AIA, we may tentatively draw the following conclusions.

Effectiveness of Notifications:

Based on our results, we conclude that notifying individuals of their interactions with AI has an effect, albeit limited. While notifications did raise the understanding of people that they were interacting with AI, especially for the group that had previously interacted with AI powered systems, it failed to increase end users' agency and trust.

Design Considerations for Notifications:

In order to increase the likelihood of a notification to be "detected" by the user, changes to the designs of the notification could be made. In particular, notification could be made to stand out more (e.g. by giving them more bold colors or making them more intrusive). However, this will not necessarily improve user understanding, user agency or trust. Furthermore, a more intrusive design might deteriorate the user experience of the application. Therefore, if the goal of Article 52(1) is to increase individuals' trust and agency through notification, we may conclude that Article 52(1) AIA is ineffective in and of itself.

Goals of Article 52(1) AIA:

Looking at Article 52(1) and the explanatory memorandum we see that the goal of the provision is twofold:

1 inform people of the fact that they are interacting with an AI;

2 avoid manipulation.

With regard to the first goal we may conclude that a notification may contribute to this goal.

With regard to the second goal, it is unclear how a notification would contribute to avoiding manipulation. First of all, it is questionable whether an actor seeking to manipulate individuals would comply with a notification requirement in the first place. Second, a notification merely draws the attention of the user to the fact that they are interacting with AI. It does not (necessarily) provide additional information on how the AI system influences user behavior. Furthermore, as our study shows, notification has no effect on end users' sense of agency.

Prior User Experience with AI:

Given the significant effect of prior user experience with AI-powered systems on the awareness of interacting with AI, it might be worthwhile to explore how to increase user knowledge of and experience with AI. By raising the hands on experience with AI, we improve user awareness and strengthen the effects of notification. Furthermore, increased interaction with AI may lead to a decreased need for notification in the future, because it will be more obvious from the circumstances and the context of use that there is interaction with AI.

Drawbacks and Focus of Notifications:

While notifying individuals has an effect, we do caution for possible drawbacks of notifications. One of the main findings of this study indicates the existence of 'banner blindness' for AI notifications. Looking at the effects of consent fatigue, i.e. the phenomenon whereby individuals who are overwhelmed by consent requests no longer make conscious choices, additional notifications may further desensitize individuals to such disclosures / warnings.[7] The risk of click fatigue or consent fatigue, may also present itself in the context of AI notifications.[8] While further research is needed to assess the effects of banner blindness in the context of Article 52(1) AIA, it might be worthwhile to explore whether the notification requirements

could be more focused on those areas where the impact of AI and the risk of manipulation are highest. This could for instance be high-risk AI systems and/or AI systems used for influencing user behavior through e.g. nudging. By reducing the number of notifications, the effects of banner blindness may be reduced and the impact of a notification may be strengthened.

Clarification of "Interacting with an AI System":

A final recommendation is to clarify the notion of 'interacting with an AI system' in Article 52(1) AIA. A large group of participants seemed to notice a difference in meaning between the use of, and the interaction with AI: in about 17.3% of cases, participants with a strong understanding of using AI did not assume to interact with AI in the applications presented. This has implications for how we inform individuals about AI, and need to consider the different interpretations "AI-powered" may bring to mind for a reader - perhaps as "use" is a one way direction term (as is "power"), while "interact" brings to mind more relational two way direction of both parties.

These policy implications and recommendations are intended to provide insights for policymakers in addressing the effectiveness of AI notifications and promoting informed user interactions with AI systems. Further research and user-centered approaches are crucial in refining and optimizing the implementation of notification requirements in line with the goals of Article 52(1) AIA.

Endnotes

1   In this report, we use the term natural persons, end-users and individuals interchangeably.

2   AIA Explanatory Memorandum, p. 14. European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. COM/2021/206 final.

3   See https://openloop.org/programs/open-loop-eu-ai-act-program/

4   This research question is in turn broken down into two sub-research questions: "How do participants notice and interpret different types of notifications of AI interaction?" and "To what extent do participants understand that they are interacting with an AI?".

5   This research question is in turn broken down into two sub-research questions: "Does the notification increase participants' sense of agency?" and "Does the notification increase participants' trust in the AI system?".

6   A sample of 660 respondents from five European countries, namely Spain, France, Germany, United Kingdom, and Sweden was collected over a period of nine days from 01.09.2022 to 09.09.2022. Three attention checks were integrated into the questionnaire to ensure high data quality. Their placement is visualized in "Questionnaire Structure". We only used responses from respondents passing all three attention checks, resulting in 469 respondents. 91 participants were from France, 94 from Germany, 97 from United Kingdom, 95 from Spain, and 92 from Sweden.

7   See e.g.: EDPB (2020) Guidelines 05/2020 on consent under Regulation 2016/679 Version 1.1, para. 87

8   Schermer, B. W., Custers, B., Van der Hof (2014), The crisis of consent: how stronger legal protection may lead to weaker consent, in: Journal of Ethics and Information Technology 16, 171-182