OPEN SOCIETY
FOUNDATIONS

AI in Journalism Challenge 2023

Final Report

David Caswell

April 2024

Table of Contents

© Open Society Foundations
Some Rights Reserved

224 West 57th Street
New York, NY 10019
P. +1 212-548-0600
opensocietyfoundations.org

The Program

Introduction

This report describes the Applied AI in Journalism Challenge (AIJC)—a competitive accelerator program intended to prototype pragmatic applications of artificial intelligence (AI) capabilities in mission-driven newsrooms around the world. The program was launched in June 2023 and ran until November 2023. Funding was provided by the Open Society Foundations. The author of this report was the lead consultant for the project.

The AIJC program was the first of its kind and therefore was itself a prototype—both of a process for accelerating the pragmatic adoption of AI within newsrooms, and also of an approach to developing investment opportunities around applying AI to news. The program was largely successful, developing a cohort of capable and motivated teams and actively engaging with them as they developed substantial new capabilities. The program also succeeded in producing a range of practical insights for its different stakeholder groups—newsrooms, program operators, investors, and the news industry generally.

This report provides a complete record of the AIJC program. The primary intention in writing it is to contribute the experience and insights accumulated during the program to the global journalism community, and to provide sufficient detail and specificity to be directly useful to anyone seeking to apply those lessons to hands-on projects, accelerator programs, and investments.

The first section of the report provides a systematic description of the program, including why it was initiated, what it was trying to achieve, how it was set up, and how it actually played out in practice. The second section reviews the tangible outcomes of the program and describes the "funneling" of projects from the proposals in the initial applications, to the teams selected to participate in the program, to the finalist projects, to the eventual program winner (and runner up). The third section lists the specific lessons learned by the AIJC community, one by one, with a detailed explanation of each lesson and the evidence for it from the program. The fourth section gathers and analyzes these lessons to provide a succinct set of specific recommendations, organized for newsrooms, for operators of similar AI in news accelerator programs, for investors, and finally for the global news ecosystem. The fifth and final section draws on those lessons to look into the near future and offers some general opinions and guidance for changes

that we might anticipate for news in the age of AI, as well as some guidance for preparing journalism for that transformation.

Context

The AIJC program took place during a unique year for innovation in journalism. In 2023, the entire global news industry collectively realized that AI was not some technical abstraction or a wooly vision of the future, but real functionality, broadly available to everyone, right here in the present. This visceral appreciation of the reality of AI was, of course, directly attributable to the launch of ChatGPT in November 2022. By the second quarter of 2023, the discussion of its capabilities and potential to cause disruptive change in journalism was intense.

This period of intense AI discussion was matched by the pace at which major new AI functionality was being released by model providers. Early 2023 was when multiple AI models clearly passed quality thresholds necessary for pragmatic application in newsrooms. The release of the GPT-4 Large Language Model (LLM) in April 2023 provided easy User Interface (UI) and Application Programming Interface (API) availability of human-like capacity for performing tasks with language—capacity that had not previously been available to even the most well-resourced news innovation teams. Similar advancements occurred in text-to-image models with the launch of MidJourney 5, in synthetic voices with the launch of ElevenLabs' "Speech Synthesis" models, with the emergence of new prompting techniques and new ways of using LLMs as "agents" to perform complex tasks, and in many other forms.

These developments set up the conditions for a new approach to AI-based innovation in news, centered on the assumption that small, low-resource teams without significant technical capacity might be able to creatively apply this emerging AI functionality to pragmatic applications in their newsrooms. This assumption stood in contrast to the "traditional" model of journalism innovation, in which large, well-resourced news organizations funded extensive "labs" populated by engineers, designers, and data scientists and backed by substantial investment over multiple years. In the spring of 2023, it was essentially unknown whether this assumption was valid, and if small teams in small news organizations could take advantage of that new opportunity. Similarly, it was also unknown how philanthropic investors like the Open Society Foundations might identify and select opportunities to enable low-resource news providers to compete and thrive in a rapidly emerging AI-mediated information ecosystem.

The AIJC program was an opportunity to explore this new approach to AI-based innovation in news while also enhancing the practical capacity of a small group of carefully chosen newsrooms to work with generative AI. Inspired by the "start-up accelerator" model of innovation pioneered in Silicon Valley by organizations like Y-Combinator, the approach combined rigorous selection with a small but material development grant supplemented by sustained coaching and expert mentoring over a fast-paced and competitive development cycle. For the participating newsrooms this provided an "immersion" in generative AI capabilities and practices, combined with sustained motivation from periodic hard deadlines and experienced coaching, all within the context of a crisply defined, pragmatic, and measurable project.

More detail about the specific components of this approach is provided below in the description of the AIJC program design and the history of its execution.

Objectives

The AIJC program had six major objectives, all intended to improve the ability of journalism to adapt to an AI-mediated information ecosystem and to facilitate a thriving and open society in an era of ubiquitous AI.

These objectives were:

- *To drive innovation*–At the most basic level the objective of the program was to discover or explore significant new ways for small newsrooms to pragmatically apply generative AI to substantially expanding their journalistic impact. At an even more basic level, the program sought to facilitate the development of valued and measurable audience or societal outcomes, achieved either directly via new audience experiences or indirectly via material improvements in newsgathering, workflows, distribution, or measurement.
- *To promote ambition*–A related objective was to find particularly ambitious and impactful applications of generative AI in news by exploiting the tension between pragmatism and transformational potential in the selection of projects and the guidance of teams. The program was explicitly not intended to pursue incremental innovations, but instead to match the transformative functionality available from large language models and other generative AI tools with equally transformative pragmatic application of that functionality within newsrooms.

- *To build innovation capacity*–An additional related objective was to develop and enhance the overall capacity of participating newsrooms to continually apply generative AI in different ways to their ongoing operations as a routine practice, by developing their skills, experience, and confidence with AI tools and techniques.
- *To prototype an innovation process*–A higher-level objective of the program was to develop and evaluate a new "accelerator" approach to facilitating AI-based innovation in newsrooms, and to identifying opportunities for investment in AI-related initiatives and strategies with potential for significant societal impact.
- *To learn*–A major objective of the program was to learn about the application of AI to journalism through engagement in hands-on practical projects. This was a significant "learning-by-doing" opportunity for everyone engaged in the program.
- *To share outcomes and lessons*–Finally, an additional objective was to motivate others by publicly communicating the specific applications of AI to newsroom tasks, developed by the participating teams, and also the performance of the overall approach to news innovation via an accelerator-style program. This communication is intended to help motivate others to pursue similarly ambitious exploration and discovery. This report is intended to help meet that objective.

The urgency of these six objectives in mid-2023 was, clearly, set by the extreme uncertainty about the potential disruptive impact that generative AI and Large Language Models might have on journalism, as well as on the larger potential risk that a new AI-mediated information ecosystem might emerge in ways that would not promote or support open, transparent, and empowered societies.

Design

The design of the program was informed by an assessment of the successes and failures of earlier initiatives of other organizations seeking to develop and facilitate innovation in journalism, and was particularly inspired by the relative success of "accelerator" models in facilitating the rapid development of valuable innovation-led start-ups.

Specifically, the program design was comprised of six distinct phases:

- **A planning phase** in which the operational details of the program were specified and prepared.
- **An application phase** centered on an open competition in which short applications were solicitated, gathered and scored, and 12 participating teams selected.
- **An education phase** centered on a series of Zoom presentations, including a structured six-module workshop, and mentoring support to the 12 teams in defining their projects.
- **A build phase** in which teams developed and tested their projects, supported by one-to-one mentoring, and competed to be one of 5 finalist teams.
- **A conclusion phase** in which the 5 finalist teams polished and tested their projects and competed at an in-person judging event to determine a program winner.
- **A documentation phase** in which the contributing projects and full end-to-end program was assessed and documented, and for which this report is the primary outcome.

Key elements of an accelerator-style approach that were present in the design of the AIJC program were:

- **Competition**. A competitive process based on evaluation against explicit criteria, both for acceptance into the program and for progression through the program.
- **Funding**. Seed funding sufficient to cover all reasonable expenses associated with the project, costs associated with pitching and preparation, and travel to the final judging session.
- **Pace**. A fast-paced structured timeline, punctuated by multiple pitch events at which progress was shared with the full program cohort.

- **Education**. A comprehensive portfolio of education and training experiences intended to inspire and empower the participating teams as they defined their projects.
- **Mentoring**. One-to-one mentoring and assistance with project definition, project execution, problem-solving andconfidence-building.
- **Ambition**. An explicit ambition to produce projects that were potentially transformative while simultaneously immediately and pragmatically valuable to the participant newsrooms.
- **Measurability**. An explicit focus on measurable outcomes directly related to impact with audiences or newsroom stakeholders.
- **Community**. Channels for communication and networking between teams, including shared meetings, a dedicated Telegram channel and an in-person gathering of the 5 finalist teams.

The intention of this design was, in summary, to assemble a cohort of ambitious, motivated, and credible teams, provide them with the resources needed to achieve ambitious objectives, have them set and commit to those objectives in clear, time-bound, terms, and then support them individually with expert guidance as they worked to achieve them. This design aligned well with the six major objectives for the program.

Execution

Competition and Selection

Planning for the program was done in May and June 2023, and followed an iterative process in which successively more detailed proposals and plans were developed. The major elements of this process included the schedule, the criteria and methodology for selecting participants, the partner consultants and organizations, and the components of the participants' experience.

The application process opened on June 20 and closed on June 30 Applicants applied online via a short application form. Selected teams would receive $5,000 development grants, finalists would receive fully funded trips to Chiang Mai for the final judging event, and the final winner would receive a $25,000 grant. The challenge was actively promoted via social media and through various organizations in contact with digital-first independent newsrooms around the world. One-hundred-thirteen applications were received. These applications were scored independently by six people from the Open Society Foundations based primarily on the application's pragmatism, potential for transformation, and measurability. The twelve program participants were selected using the combined scoring from all six evaluations of all applicants, adjusted slightly based on a desire to form a cohort that was diverse in terms of project goals and geography. This assessment required multiple meetings and extensive follow-up, and the winning participants were notified on July 11, 2023.

Mentoring

Expert mentoring was a pivotal aspect of the AIJC program. The mentoring provided to the AIJC teams consisted almost entirely of one-to- one engagement with each team, usually on a weekly basis but sometimes more frequently if special topics needed to be discussed or special situations arose. The mentoring generally was founded on both practical knowledge such as models, other tools, techniques, people to talk to, and approaches to consider, as well as on providing a framework with which to think about projects and approach their definition, design, and development.

The two mentors for the program were:

- David Caswell: Lead consultant on this project.

- Bahareh Heravi: Professor of AI & Media, Institute for People-Centred AI, University of Surrey.

Education

The program kicked off with a two-hour introductory session on July 20, held via Zoom, providing a broad overview of Generative AI and its potential application in newsrooms, including numerous specific examples of that potential. This was closely followed by one-to-one mentoring sessions with each team focused on defining or clarifying each project. Participating teams then attended a special version of the London School of Economics and Political Science's Journalism AI Academy for Small Newsrooms, comprised of six modules presented by guest speakers via 90 -minute Zoom sessions.

These modules were:

- **Module 1: What is AI and how is it used in journalism?** presented by Joanne Kuai from Karlstad University (August 7).
- **Module 2: Tools & Applications,** presented by Gary Rogers, Senior Newsroom Strategy Consultant at Fathm and founder of the Press Association's RADAR automated newsroom (August 9).
- **Module 3: Data & Technology,** presented by Jonathan Soma, Knight Chair in Data Journalism at Columbia University (August 11).
- **Module 4: People & Culture**, presented by Uli Köppen, Head of the AI + Automation Lab at Bayerischer Rundfunk (August 14).
- **Module 5: Risks & Concerns,** presented by Tess Jeffers, Director of Data Science at the *Wall Street Journal,* and Alyssa Zeisler, Vice President of Product Management at Dow Jones (August 16).
- **Module 6: From Ideation to Implementation**, presented by David Caswell, lead consultant for the AIJC program (August 18).

In addition to these educational sessions, this period was also focused on working with the participant teams to help them define and develop their specific project concepts, including via one-to-one mentoring sessions. All teams then presented their proposed project designs to the full cohort in an extended Zoom presentation held on August 17. A final presentation on AI ethics and copyright issues was presented to the full cohort on August 23 by Nick Diakopoulos, then on a research sabbatical at the University of Amsterdam studying generative AI.

Build

By this stage of the program the participating teams were quite well prepared for project development. They had been immersed as teams in AI concepts and techniques for a month, had produced clear and specific project designs, had established relationships with their mentors, had received their $5,000 development grants and were motivated by the opportunity to participate in the final competition in Chiang Mai. They then accelerated their project development, building out their designs while supported by one-to-one mentoring focused on the specific needs of each individual project. This mentoring took the form of weekly 30-minute meetings, supported by email exchanges, and less frequent longer sessions for special situations. The level of engagement varied somewhat from team to team and the weekly meetings were variously used for progress reports, general advice, troubleshooting specific issues, evaluating options, instilling confidence, and preparing for the competitive evaluation. The only cohort-wide session during this phase was a short "Pitch Preparation Workshop," presented by Alan Soon and Rishad Patel from Splice Media on October 3, and the judging session to select the finalists, which took place on October 12.

Judging

The judging session was conducted on a two-hour Zoom video call with all 12 teams, who pitched their projects to five independent judges in fast-paced 4-minute presentations followed by a few minutes of Q & A. This session was preceded by detailed practice and feedback sessions from the program mentors, but the mentors were not involved in the judging itself. The five judges were selected for their deep familiarity with the issues involved with AI in journalism, for their familiarity with independent mission-driven news, and their standing in the global journalism community.

These five judges were:

- **Gina Chua**: Executive editor of Semafor and previously executive editor at Reuters, editor-in-chief of the *South China Morning Post* and editor of the *Wall Street Journal*'s Asian edition.
- **Abhijit Das:** Programme director at Stichting Democratie en Media foundation, and co-chair of the Journalism Funders Forum. Abhijit served as head judge and coordinated the judging activities for AIJC.
- **Valer Kot**: A senior media advisor at Media Development Investment Fund (MDIF) and previously chief operating officer at Piano Media and director of e-publishing at SME in Slovakia.

- **Marina Walker Guevara:** Executive editor at the Pulitzer Center and previously deputy director of the International Consortium of Investigative Journalists (ICIJ) and a 2018-2019 John S. Knight fellow at Stanford University, focusing on AI in journalism.
- **Sabrina Argoub:** Programme manager of JournalismAI at the London School of Economics and Political Science (LSE), and leader of their AI Academy for Small Newsrooms program.

The open judging session was attended by all participant teams, the AIJC program team, academic and industry observers, and the five independent judges. The session was followed by several closed deliberation sessions among the judges, with the mentors available to answer specific questions from the judges. Judging proved to be difficult for many projects due to compelling projects from more than five participating teams, and it took several days to select the five teams that would progress to the final stage of the program.

Following selection of the five finalists, the program mentors met with most of the unsuccessful teams to provide feedback and to suggest options for furthering their projects. The successful teams returned to their development work, improving and "polishing" their projects, including testing and evaluation of audience and journalist responses. During this period the one-to-one mentoring of the five finalist teams continued as before and preparation for the final judging event was completed.

The final judging event was held in-person at the Splice Beta journalism festival, held in Chiang Mai, Thailand, on Thursday, November 9, 2023. Prior to traveling to Chiang Mai each finalist team was asked to produce a three-minute video presentation of their project, which were then combined into a "demo reel" of the finalist projects (https://vimeo.com/881894897). Two representatives from each finalist team attended, as well as the two program mentors, members of the program management team, and three of the previous five independent judges—Gina Chua, Valer Kot, and Marina Walker Guevara.

The judges were by this point well prepared to assess the nuances of each project because they had already been through the finalist selection process and they had seen the demo reel videos. The judging event was a 90-minute closed session in which the three judges received approximately 10-minute presentations from each of the five finalist teams, and posed questions to each team. The judges then held a closed deliberation session for several hours. The program mentors attended the deliberation session to answer specific questions and to take notes, but otherwise did not participate in the deliberation conversation. Again, judging was difficult

due to the presence of multiple compelling projects. The judges chose Rappler's TL;DR project as the winner, but also chose to give Cuestión Pública's Odin project a formal honorable mention. The decision was announced publicaly in an hour-long AIJC event held on the main stage at Splice Beta that evening.

The Splice Beta event marked the end of the active part of the AIJC program. There were some follow-up meetings with specific teams and program managers in the following weeks, including a debriefing or "retro" session reviewing the entire program. Attention then turned to documenting the program and its lessons and the production of this report.

Outcomes

AIJC produced two kinds of outcomes. The most valuable was the experience, knowledge, insight, and lessons accrued by the participants and the organizers over the full arc of the program—the rich but general benefits that will hopefully continue to empower the organizations and individuals who participated, that will hopefully contribute towards a healthier information environment in a world of ubiquitous AI..

But in order to achieve those generalized outcomes we first needed to achieve a series of very specific, tactical, step-by-step results in both the execution of the program and in the definition and execution of each of the 12 individual projects in each team. Understanding these specific outcomes—especially the individual projects—will be useful to those seeking to understand the program end-to-end, wishing to get a more detailed sense of how the lessons and recommendations described in this report were obtained, or just looking for more nuance. This section describes those specific outcomes in the order in which they were achieved, beginning with the initial application process and the selection of the participants, the definition and building of those projects, the selection of the finalists, and the concluding selection of a winning project, as well as an "honourable mention."

113 Applications

The AI in Journalism Challenge began with an open solicitation. Applicants were intentionally asked for relatively little information in order to encourage applications from less resourced or busier newsrooms. The primary information required was a response to the question: "*What business or journalistic impact would you like to make through this program?*" and a response to the question: "*How might you measure that impact?*" We also requested some minor administrative data including the size of the organization and the numbers of hours per week the applicant team was prepared to contribute to their project. The solicitation was open for two weeks and was marketed via social media posts and through journalism-related organizations. We attempted to ensure that the marketing effort reached prospective applicants globally and was not overly focused on the major media markets of Europe and North America. The marketing effort was also disproportionately targeted to mission-driven news organizations rather than commercial or openly partisan news organizations. We also stressed that we were very open to "no code" projects undertaken by non-technical staff, as well as to solutions developed with software engineers

or data scientists. By the close of the application period, we had received 113 applications.

Those 113 applications can be grouped into some rough categories. A small portion of applications were either too brief or too vague to get a good sense of what the applicant had in mind. A larger portion of applications showed general excitement by the prospect of applying AI in their newsrooms, but were vague or uncertain about what they might apply it to. A few applicants had very specific project descriptions that revealed that they had been thinking about applying AI in their newsrooms for a while. The largest group of applications had a general idea of an opportunity or "pain point" that they wished to address with AI, and had a strong sense that AI could help, but were clearly new to AI and unsure about the specifics of applying AI tools to their task. This group was disproportionately represented in the selected projects.

The applications of AI specifically proposed by the initial applicants was quite diverse, however some general categories were apparent. Some were focused on optimizing or automating existing tasks within newsrooms. Some were focused on using AI to reformat content or to move content to a different medium. Some were focused on assessing the impact of journalism in delivering on a particular social or democratic mission. Some were focused on automating newsgathering by reading webpages and documents at great scale. Some focused on surprisingly intricate objectives involving expanding the use of data journalism projects or archives.

We were surprised by how few applications we received from the large media markets of Europe and North America, even when taking into account our marketing focus. Our concern that our applicant pool might be somewhat biased towards these markets proved unfounded.

12 Projects

The process of selecting 12 participants from the 113 applicants, as described in the project execution section above, was based on independent scoring of each application against a written set of criteria by six individuals from the programme management team and the sponsoring team at the Open Society Foundations. This independent scoring was then followed by group discussion to review tiebreakers and to discuss a few ambiguous issues, such as concerns raised in the application text, unexpected low "committed hours" for the applying team or challenges in distributing a development grant in some countries. The independent scoring was

generally consistent between individuals, and the final identification of the 12 "winning" participants was relatively straightforward.

There were three primary criteria for selection of program participants based on the description of the impact the applicant sought to achieve and the statement about how they intended to measure that impact. These criteria were pragmatism, transformational potential, and measurability.

- *Pragmatism* was assessed in terms of potential to actually deploy the AI solution developed in the program in routine day-to-day operation withing the news organization. This was not just about technical feasibility but also about the overall likelihood that an applicant's ambition could be achieved given their size, experience, and resources.
- Transformational potential was assessed in terms of ambition to fundamentally change and improve how the applicant news organization operated in some way. We sought applicants that were interested in applying AI in ways that were equivalent to the transformational potential of the underlying AI and LLM technology that would be used in the project.
- Measurability was assessed in terms of the willingness of the applicant to approach their project based on their specific impact objective, and their clarity on how to assess their success in terms of that objective. These criteria were not just narrowly about data-driven development, but also about assessing a mindset of measurement and accountability to audiences.

The intention in using these criteria was to exploit the obvious tension between pragmatism and transformational potential to produce projects and outcomes that were disproportionally likely to deliver genuinely significant improvements to the participating newsrooms and to others.

In addition to these three primary criteria there were also secondary criteria focused on the assessed commitment of the applicant, including the stated hours-per-week that an applying team was prepared to commit to the project. This selection process produced 12 program participants, and acceptance letters were issued on July 11, 2023.

The selected participants were, in random order, as follows:

1. Agência Pública (Impact tracker, audio article reader)

The core Agência Pública team includes a skilled developer and an editorial lead, and they developed two separate projects during their AIJC program involvement. Their primary project is an LLM-enabled impact tracker that will replace their existing keyword-based impact tracker tool and will enable faster, more accurate, and more comprehensive monitoring of the impact of their journalism across their 800 publishers and the wider Brazilian media ecosystem. Their secondary project is a workflow to convert their article journalism into an audio product read by a synthetic voice. The new impact tracker is based on a series of API calls to OpenAI language models. The tool uses the low cost/low quality GPT-3.5 model to identify mentions or references to Agência Pública journalism in a large corpus of digital media sites, then uses the GPT-4 model to interpret the specific kind of impact (kind, scale, etc). Much of the work on this tool has been in developing prompts for each model that can reliably deliver the required accuracy. The secondary audio project is intended to expand the product range that Agência Pública offers, and thereby expand the audience for its journalism. It was based on a synthetic voice trained (with permission, using tools from ElevenLabs) from the voice of a well-known Brazilian broadcast journalist, and was centered on an audio-specific workflow that provided human editing of a GPT-4 generated script for high quality audio experiences.

2. Raseef22 (A search engine optimization workflow)

This team was enthusiastic in their ambitions for using AI, even though they had no previous technical experience with AI or product development. Following some initial exploration of a chatbot project they pivoted to an LLM-based search engine optimization (SEO) project. A lot of their work has been in developing prompts for GPT-4 (including considerable work on few-shot learning based on Google's SEO documentation). They essentially assembled a repeatable test-driven prompt pipeline from design, iteration, evaluation to A/B testing of output with audiences—with SEO as the first exercise of that pipeline. Their SEO workflow uses LLMs to identify keywords, write descriptions, and rewrite or suggest re-writing of some texts. The "closed-loop" from A/B testing back to the iterative prompt design was intended to continually optimize their SEO performance over time.

3. The Initium (A Chinese language monitoring tool)

This team included technical developers and data engineers, however these members were difficult to engage with directly due to a language barrier, and our primary contact was with a product manager who represented the full team. Their project was a tool to continually monitor Chinese language media globally, providing stories and contextual background based on input words, phrases, and sentences. They have a well-defined later ambition to adapt this tool to "read between the lines" of content from inside the People's Republic of China , extracting nuance and interpretation that gets through the censorship filter. They were also interested in an architecture that roughly uses the same tool to listen to the news environment and also listen to audiences (instead of input keywords, etc), with their newsroom acting as a kind of "editorial switchboard" in the middle. This was an impressively ambitious project, however the tool that they built during the program itself was, unfortunately, somewhat difficult to evaluate for non-technical judges because it was deployed entirely within the Google Docs environment—with inputs via Google sheets. This made sense as a workflow choice, but also made it difficult to evaluate the full scope of the project.

4. The Conversation ("Microsites" and reversioned content)

This project originated with The Conversation Indonesia, with participation from The Conversation's parent organization in Australia and in other locations. The team included an editor, a developer, and a product manager. Their project was to develop an entirely new channel for their traditional journalistic output (which is expert-written articles) that communicated their journalism in new text and graphical formats that are more suited to younger, less educated audiences. The project used AI/LLMs to do the reversioning, and they sought to publish the re-versioned content on a series of "microsites"initially as an experiment focused on the upcoming Indonesian elections. They built a comprehensive prompt library for the new formats and used social media to do early tests of these formats. They also put considerable work into the editing cycle for this new content, with a firm commitment to maintaining their reputation for and commitment to authoritative, expert-certified journalism in the new AI re-versioned formats.

5. Cuestión Pública (Story context from structured data)

This was a small, thoughtful team comprising of the chief executive officer and data editor from Cuestión Pública, supported by a consulting company with expertise in data engineering. The project essentially aimed to convert structured

data about elections, political parties, powerful people, and companies (which Cuestión Pública had already assembled for a series of "game" interactives and other data-driven news products) into a series of shorter, more accessible stories that could be distributed via social media. They were particularly focused on enabling the fast production of accessible stories based on recent or breaking news but grounded and contextualized in their structured dataset. They took two parallel approaches to this project—a shorter-term approach that used GPT-4 alone to directly convert the structured data into social media posts, and a longer-term approach based on developing an embeddings corpus from their existing content and structured data, held in Elasticsearch, and performing "lookups" against this embeddings corpus for use in producing the social posts.

6. Zamaneh Media (Newsletter production)

The core Zamaneh team was a news product lead and a business development lead. Their project has developed a tool to reduce the effort required to produce email newsletters from their content (often long, philosophical pieces about the Iranian situation, written for an educated, academic, and influential audience), and also to begin improving the quality and accessibility of their newsletters. Much of their work during the program was on developing a process for specifying, designing, evaluating, deploying, and measuring the use of prompts, and also on evaluating the suitability of GPT-4 for language tasks in Persian. Towards the end of the program they developed a compelling no-code user interface to express their carefully developed prompts to their journalists.

7. Rappler (New brand using AI reversioned content app)

The Rappler project has a substantial team with leadership from their editorial and product organizations. Their project is to set up an entirely separate brand (TL:DR) aimed at reaching younger audiences in the Philippines and globally, using AI reversioning of Rappler's existing journalism output. The new brand focuses on softer news and on communicating in short posts, graphical stories, and video. The workflow to support this will be operated by the existing editorial workforce. The Rappler team had a very comprehensive business plan for this project and full support from their senior leadership, and they had already begun publishing summaries using the adapted text workflow early in the program. They validated their design for their semiautomated text-to-video workflow and also developed a "news-as-comics" workflow using text-to-image models—in partnership with their existing illustration staff. Rappler also worked to evaluate

the OpenAi models in different Filipino languages, with the intent of supporting four languages in their new, reversioned formats.

8. PumaPodcast (Workflow efficiencies for news bulletins)

This team came to the program with a well-formed vision for their project, which was focused on an improved podcast production workflow rather than on a direct audience-facing objective. Their goal to improve the efficiency of all the multiple, chained tasks surrounding their core newsgathering and audio presentation functions, such as story generation, script writing, show notes, social promotions, and clip selection. Interestingly, this project did not include synthetic voice generation, both because professional audio presentation is a differentiator for them and also because professional human voice readings are a relatively easy and inexpensive task for them.

9. Scrolla (A multimedia workflow using AI reversioning)

This was a small, scrappy team, with a growing audience and a strong desire to make their content more accessible to more people in South Africa and Nigeria and with ambitions to expand to Kenya. They worked on a set of inter-related projects during the program, in two categories. In the operational category, the project includes content creation for an African audience such as entertainment stories, localized and local vernacular weather reports, internal editorial assistance with headlines, SEO, social posts, and some small experiments with an “agony aunt” column (which answered agony-aunt style questions from audiences), and a game. The second category is in using AI to train their journalists—a broad range of people from the covered communities who report and write part-time and who often have linguistic and educational barriers to expression. Scrolla developed approaches to using language models as an integral part of their training and tooling to enable this journalistic workforce to effectively create publishable material from on the ground, local reporting.

10. Rubryka (Contextualizing stories from archives)

This team includes a data journalist and an editor from Rubryka’s English version. The project was originally aimed at extracting descriptions of Russian war crimes in Ukraine from their archive, for the purpose of generating 3D graphical representations of the crimes. As the project progressed, this was modified to a simpler goal of generating social media posts that use stories from archives related to current events, locations of breaking news, and individuals in

breaking news.The approach was based on a sophisticated no-code or low-code solution that included use of an LLM to extract structured information from archives, stored in an AirTable database and managed via an online process automation tool (Make.com). The initial use case—social media posts—was an early, practical use of this workflow but their ambition was to use it in the future to extract material from archives in structured form to use in multiple ways, including for graphical or map-based depictions and the original intent of 3-D visualizations.

11. Daraj (Impact tracker)

This was a non-technical team, without prior exposure to AI, but they put together a thoughtful objective and prompt-based approach to achieve it—a semiautomated impact tracker based on ProPublica's impact tracking process and intended to replace an existing manual workflow. They put considerable work into evaluating the effectiveness of GPT-4 in Arabic, and also into the careful design of their impact-tracking workflow. This work includes a careful calculation of expected time savings and cost savings from the new workflow—easily the most systematic approach to documenting efficiency savings that we saw during the program. As a non-technical team, they were dependent on subscriptions to 3rd party tools, like Dataminr, to implement substantial parts of their workflow.

12. Meduza (A summarization pipeline)

This was a highly engaged hands-on team with some technical ability led by an editor. The traditional output of their newsroom is very long (10,000 words) and highly nuanced pieces in Russian, aimed at an educated, influential audience. These pieces have limited reach and so the team's project is to develop versions of their journalism for a broad non-expert audience in English. They have carefully evaluated techniques and prompts and built a GPT-4 enabled, multi-prompt, summarization pipeline, including production of text-to-image prompts for illustration. They are also working on two user interfaces for this pipeline—a simple one for routine use and a more complex one for more advanced uses. They spent considerable time evaluating fine-tuning as an option for maintaining the essence and nuance of their articles and editorial voice in their summaries.

5 Finalists

The finalist teams of the AIJC were selected in a judging session held on October 12, 2023, when all 12 participating teams presented their projects. As described earlier in this report, the presentation session was conducted on a two-hour video call attended by all participant teams, the AIJC program team, some observers, and five independent judges—Gina Chua, Abhijit Das, Valer Kot, Marina Walker Guevara, and Sabrina Argoub. Prior to the session, the judges were provided with a written description of each project for orientation, and with a set of three judging criteria, as follows:

- *Practicality*: Can the project be adopted into routine projection beyond a one-off prototype? Is the project measurable? Is it cost-effective? Is the project limited in application in some way?
- *Transformational potential*: Does the project transform the organization's newsroom? Could the project transform the news industry? Is the project scalable?
- *Imaginativeness*: Is the project unique? Does it add new value to existing ideas? Does it expand the concept of what journalism might become using AI?

These criteria were modified from the original participant selection criteria, with "Imaginativeness" being substituted for "measurability."

During the judging session each team was allocated four minutes to present their project, followed by a few minutes of questions and answers. Each team had been provided with a one-to-one timed practice session on the day before the judging event, and the teams generally kept within their allotted times. Following the session, the judges conferred in several video calls and via email, including asking specific questions about the projects from the program managers. Making final selections proved to be quite difficult and required considerable discussion among the judges.

The following descriptions of the selected finalists highlight some of the characteristics of these projects that appealed to the judges during this process based on notes from the deliberation conversation. These notes did not repeat detailed descriptions of each project, which had been provided earlier.

1. Agência Pública–from efficiency to reinvention

The judges found this project to be practical, transformative, and imaginative.

The impact tracker was seen as immediately practical because it partially automated an existing workflow that consumed a considerable amount of the time of skilled journalists, and because it provided an immediately applicable and immediate benefit to the newsroom. The burden of impact tracking was intimately familiar to Agência Pública's journalists as necessary but journalistically unproductive work, and so the benefits of automating that work were obvious.

The impact tracker was also seen as potentially transformative because of Agência Pública's expansive view of it as infrastructure for a much larger AI-enabled newsgathering system, capable of continually monitoring a large number of websites. A near-term ambition for this extension into newsgathering was the potential to continuously monitor more than 800 websites of evangelical Christian organizations throughout Brazil, using AI automation controlled by and reporting to a single journalist as a part-time task. The implementation of this ambitious objective could be done using the same infrastructure and prompts developed for the impact tracker.

This project was also seen by several of the judges as imaginative, not just for the use cases themselves but also for the strategic path that they provided from immediately practical to radically inventive. The second project—which used GPT-4 to create scripts for audio experiences from text articles and used synthetic voices to render those scripts into publishable audio products—was also seen as both meeting an immediate need and also taking a step towards a longer, much more impactful vision.

2. Cuestión Pública –integrating news and structured data

Cuestión Pública's project, Odin, was seen by the judges as potentially transformative, not only for the immediate application of automated contextualizing of breaking news based on Cuestión Pública's comprehensive dataset of Colombian political and business elites, but also for the potential that it revealed for using high-quality structured datasets maintained by journalists as a basis for customized stories.

The practicality of the Odin workflow and application in publishing was more difficult for the judges to assess than some other projects, because of its more technical "retrieval-augmented generation" (RAG) design. Prototyped examples were presented during the judging session, however it was Cuestión Pública's history of delivering and publishing complex news products from data, such as their "Game of Votes" interactive experience, their deep familiarity with the underlying dataset upon which Odin was based, and their practical vision for the project that satisfied the judges that this was a thoroughly pragmatic approach.

The transformative potential of the Odin project, for both Cuestión Pública's mission and for news production more broadly, was the primary consideration of the judges during the deliberation session. The concept of creating and maintaining a factual record of the structure and relationship of societal elites, and then automatically using that to instantly contextualize breaking news was seen as almost revolutionary. There was discussion of how this approach might work on a broader scale, or within the U.S. context, or in other domains, and it was clear that this was a concept that genuinely excited most of the judges.

The imaginativeness of the Cuestión Pública project was also admired by the judges. They saw it as a clear example of an attempt to use LLMs to fundamentally expand the concept of what journalism could become in an AI-mediated information ecosystem, in terms of its coverage, speed, and accessibility.

3. The Conversation–enabling the long tail of news

The Conversation's "micro-site" project was an excellent strategic fit with their mission of publishing high-quality content written by leading experts and academics. The judges felt that although topic-based "microsites" had proven difficult for publishers to create and maintain at scale in the past, using manual workflows, the new potential for LLMs to both populate these sites and keep them updated represented a fresh approach.

The discussion of the practically of The Conversation's "microsite" project was dominated by the views of several of the judges about the usefulness of microsites as a concept, and the risk of a combination of low traffic and high maintenance burden. This discussion concluded with an assessment that the potential of LLMs and other AI models to help create, maintain, and scale the use of microsites might make them more practical as a channel then they have proven to be using fully manual publishing. The immediate focus of the prototype microsites on the Indonesian election was seen as clearly practical and valuable.

The discussion about the transformative potential of AI-assisted microsites centered, essentially, on the potential for extreme personalization of topic-level production and curation using LLMs. This potential was seen by one judge as being a particularly good fit with The Conversation's expert-based publishing model, but there were questions about how well it might apply to other, more general, forms of news publishing.

In a similar way the imaginativeness of The Conversation's project was seen as closely tied to their expert-based publishing model. It was pointed out that a "long-tail" of microsites, producing an extremely granular categorization of news, might also fit well with other forms of publishing, such as scientific publishing.

4. Rappler–going all-in with AI formats and media

The judges were impressed by Rappler's simple solution to a key challenge facing any news publisher using AI to render news content into different formats and media—setting up an entirely separate brand for separate content targeted at a separate audience. This practical and relatively low-cost approach to serving AI "re-mediated" content to target audiences demonstrated that complex, lengthy development projects are not always needed to execute ambitious strategies.

The business-focused practicality of Rappler's TL;DR project, combined with its ambition, impressed the judges. The clear strategic goal of serving younger audiences using the new capabilities of AI was admired, as was the clear and relatively straightforward plan to do that with an entirely new brand.

The transformative potential of the Rappler project was seen in its clear demonstration of the potential of AI-enabled content reversion to address large systemic challenges in news publishing. It was seen by the judges as a straightforward and unsentimental application of these new capabilities to well-understood but intractable business challenges.

The judges were also impressed by the imagination and range of the formats and media that Rappler chose to convert their content into using partly automated workflows, which not only included text formats, such as summaries and translations, but also graphical formats (comics) and AI-produced video formats. It was mentioned that the decision to pursue these AI-enabled content re-versioning objectives was ambitious but also clearly audience-driven.

5. Zamaneh–building a tool while building enduring capacity

Although Zamaneh's project was perhaps not as sophisticated as some other AIJC projects, the judges were very impressed by the systematic and professional way

in which this low-resource team was able to build a practical tool that addressed a specific challenge in their newsroom (Newsletter Hero) while also developing robust AI and prompting capability that could be applied generally, to many other challenges. This journey was seen as an excellent practical demonstration of the accessibility of professional AI capabilities to smaller newsrooms.

The Zamaneh team were relentlessly practical in their design and implementation of Newsletter Hero because they were focused on replicating their existing well-understood newsletter production process using an LLM-based process. Their approach of studying the existing process carefully, assessing the key AI-provided functionality carefully (including translation) and implementing a tool that could be immediately useful was appreciated by the judges.

The judges saw the transformative potential of not only the Newsletter Hero tool itself, but of the process that produced it and the potential of that process to be used more extensively within Zamaneh and other newsrooms. The ability of the Zamaneh team to develop competence and confidence in all aspects of applying AI to newsletters within a few months—from assessment to requirements gathering to prompting to implementation and testing—was seen as a demonstration of the potential to apply these tools in any newsroom.

The Zamaneh project was referred to by one judge as "quietly imaginative," meaning that imagination was apparent in the step-by-step implementation details of the project rather than in the overall ambition. The Zamaneh team's description of their ambitions to continue and extend their newsletter portfolio using Newsletter Hero and to extend their approach to automation to other workflows in their newsroom was also appreciated by the judges.

1 Winner: Rappler's TL;DR

The selection of an overall winner of the AIJC program occurred at the Splice Beta journalism conference, held in Chiang Mai, Thailand, on November 9, 2023. As earlier described in the program execution section of this report, three judges—Gina Chua, Valer Kot, and Marina Walker Guevara—received presentations from the five finalist teams. They then held a closed deliberation session for several hours and chose Rappler's TL;DR project as the winner.

The judges assessed the Rappler project largely from a business and potential impact point of view. The primary business goal of the project—to expand Rappler's journalism to younger audiences who were less comfortable with long text articles—was a familiar challenge to them, and Rappler's use of AI to reversion their journalism into new formats and media seemed to be a plausible solution. The ambition and comprehensiveness of this solution was also admired by the judges, including the range and complexity of reversioned outputs (text, comics, video) and the use of automated translation to produce content in multiple Filipino languages.

The judges also spoke about the practical aspects of the TL;DR project, especially about the use of an entirely separate brand as a simple way of getting around the challenge of distributing non-traditional forms of content to new audiences while simultaneously continuing to serve their existing audience with familiar formats and styles.

The aggressive execution of the project was also a factor, including the live publishing of AI-generated summaries on most Rappler content, and the decision to publish unedited AI-generated summaries with clear and unambiguous labeling ("*This is AI generated summarization, which may have errors. For context, always refer to the full article.*"). This clearly demonstrated a willingness to fully engage with the potential of AI to address real business issues, without hesitation or half-measures.

The way in which Rappler went about their TL;DR project also seemed to make an impression on the judges. Rappler did not seek to replace or reduce their existing staff in executing TL;DR. Instead, they sought to develop the workflows for this new brand entirely within their existing newsroom, rather than within a separate, dedicated team. They also meaningfully involved existing staff in the production of new content using the new AI workflows, for example in empowering their existing illustrators to produce the new AI-generated comic content to quality standards that they required.

Overall, the Rappler project seemed to provide a template for how a relatively large news organization, with a respectable reputation for journalistic integrity but with relatively few technical or financial resources, could decisively use AI in a range of ways to directly address some of its most fundamental and intractable business challenges. Of all of the AIJC projects, Rappler's TL;DR probably came the closest to showing how some of the most threatened news organizations globally—mid-sized regional or metro publishers—might productively and pragmatically use AI to extend the reach and impact of their work.

1 Honorable Mention: Cuestión Pública's Odin

The AIJC program had been intended to produce a single winner, however the judges were struck by the ambition and long-term potential of the Cuestión Pública Odin project and wanted to explicitly recognize that achievement.

In their discussion, the judges focused on the potential of Cuestión Pública's approach to eventually reshape and expand how large portions of journalism might be done. They saw this as a glimpse of how journalism might someday employ LLMs to construct and maintain large knowledge bases of journalistic content, as Cuestión Pública had done manually with its structured dataset of Colombian political and business elites, and then use those knowledge bases to provide immediate context to new developments in accessible ways, just as Cuestión Pública was doing in its creation of social media posts.

The judges were also impressed by the focus that Cuestión Pública had on maintaining their unique and distinctive "editorial voice" in the Odin project, to the extent of considering the fine-tuning of an LLM to maintain that unique voice. This unabashed ambition to fully embrace both the journalistic potential of LLMs as well as the differentiating value of editorial voice was referred to several times in the discussion.

Finally, the judges appreciated the technical ambition of the Cuestión Pública project, including its exploration of a Retrieval Augmented Generation (RAG) architecture and of fine-tuning.

The Odin project was not yet scalable, and still relied on manually assembled structured data, and its near-term reach and impact were still uncertain. However, it did provide a vision for a much more radical interpretation of what journalism could become in an information ecosystem empowered by AI—delivering new journalistic impact in an intentional and differentiated way, using the full range of available tools.

Lessons

The AIJC project was relatively small—12 participants—and was not conducted with any formal research component. By design and intention, the projects were pragmatic and entrepreneurial, and no quantitative or qualitative research data was collected. Furthermore, this was the first initiative of its kind and it was designed and initiated at a time when generative AI was relatively new and its application in newsrooms highly uncertain. Nonetheless, there are many lessons that became apparent as the teams became more proficient in using AI and as their projects developed.

These lessons come from patterns observed within teams and across teams. They are either apparent from specific examples within the cohort, or from comparisons across all 12 teams. They are interpretations from a small sample, and should be treated carefully, however validation of many of them can also be found in examples from outside of the program.

A few considerations should be kept in mind when reviewing these lessons. One is that the AIJC cohort represents a selection bias towards engagement with AI, and that newsrooms that did not apply for the program, or were not selected, might have had different outcomes. Another is that these teams were doing these projects during a time when AI and its application were at the peak of their "hype cycle," and so may have been more motivated than they would have been in a less AI-saturated time.

All teams spoke English fluently, and with one partial exception, usually as a second language. While that did not seem to be a disadvantage in most cases, it does represent another bias in that these teams had ready access to both AI tools and to information about the tools that may have been less accessible to non-English speaking newsrooms.

In describing these lessons, I will aim to give specific examples, however sometimes I may speak generally to avoid unnecessarily specific implied criticism. All participants in all teams performed impressively during the course of this program, and their success greatly exceeds any minor issues they may have encountered along the way.

Finally, these lessons are inevitably biased by my own experience and views about the role of AI in journalism. They are neither complete nor objective, and others may have observed different lessons.

Opportunities to apply AI span the news value chain.

The 12 AIJC teams identified, selected, designed, and implemented AI-based projects across the full value chain of news, including newsgathering, news production, distribution, and consumption experiences, including many projects that incorporated aspects of several of these stages. This range of opportunities was also apparent in the 113 initial applications to the project, and it suggests that the potential of AI and ambitions for its useful deployment in news organizations may not be limited to just content generation.

There were several news gathering projects in the cohort, including The Initium's project, Cuestión Pública's data journalism project, the underlying infrastructure built by Agência Pública for their impact tracker and aspects of Daraj's impact tracker project. All of these projects sought to extract new news information from either websites or, in the case of Cuestión Pública, from structured datasets, and then use LLMs to analyze, synthesize or contextualize that information into either information useful to newsroom staff or into a draft of a publishable news product.

Content production projects were the largest category of application and included production of new forms of content suitable for new audiences, for example by Rappler or The Conversation, or automated support of existing content production, for example by Meduza's advanced summarization or Zamaneh Media's Newsletter Hero project.

Distribution of content was the primary focus of Raseef22's AI-assisted Search Engine Optimization project and Zamaneh's Newsletter Hero project but was also a secondary part of several other projects via the automated generation of draft social media posts, for example by the Cuestión Pública Odin project.

The provision to audiences of different consumption experiences was most dramatically demonstrated by Rappler's TL;DR project, but was also apparent in others, such as The Conversation's microsites or in the audio versions of news stories created by Agência Pública.

This diversity of projects across all parts of the news value chain suggests that AI has applicability, and therefore has the potential to change, all stages of the news value chain. It confirms the observation that AI and especially LLMs are general technologies that will be applied generally, and also that early adopters are already well aware of this broad range of potential and well able to build applications across that range.

Digital news providers can apply AI relatively quickly.

The 12 AIJC teams developed from essentially novices to quite skilled AI practitioners in just a few months. Prior to beginning the program most of the AIJC teams had minimal experience or even familiarity with generative AI, large language models and prompting. They generally had a problem area in mind and an idea that AI could help, but typically very little specificity and often no clear understanding of how the available tools might be applied in a practical and professional way to that problem or need.

In the space of about four months, most of these teams went through a quite remarkable transformation. The most noticeable changes were in the confidence that the teams exhibited in working with AI, the sophistication of their thinking and discussions about applying AI, their willingness and ability to explore, contextualize and apply learnings or tools on their own and, of course, their production of usable AI products and often testing of those products with audiences or journalists. This progress was not primarily due to training provided by the AIJC program, but by the direct and sustained hands-on engagement of the teams themselves with AI tools and the real-world problems they were attempting to solve.

A particularly illustrative example of this is the journey taken by Zamaneh Media, one of the five AIJC finalists. The two Zamaneh team members began the project with almost no background in AI or automation, no prompting skills, and no technical expertise. Over the course of the project, however, they engaged thoughtfully and earnestly with GPT-4, with their problem area (newsletter production) and with the challenges of defining and evaluating increasingly complex prompts. They developed familiarity and confidence with what LLMs could and could not do, with how prompting could control the LLM in nuances ways and with how to interpret their underlying real-world task in ways that let them apply the LLM. The result was not only a successful application of an LLM to their project task, but also an enduring comfort and familiarity with thinking through tasks and challenges in ways that let them apply LLMs productively. It is clear that this team will be capable of ongoing work in applying AI to many diverse tasks, and also that they will continue to develop and grow as prompt editors, architects, and managers of newsroom automation.

This transformation from novice to relative expert seemed to be the default experience across almost all the teams, and it suggests that the opportunity to change how news providers think about and apply AI may be easier to realize

than might be assumed. It certainly appears to be far easier, for example, than thinking about and applying the previous generation of AI, based on highly technical concepts, data science talent, and engineering resources.

Barriers to applying AI in news are relatively low.

The speed with which the AIJC teams became comfortable applying AI in their newsroom, and the success of most of them in developing practical applications came despite a number of disadvantages. The cohort represented teams that were geographically located far from the "tech scenes" of Europe and North America. Most of the teams were relatively small, with relatively precarious funding situations, and with relatively little access to skilled software developers. The funding available to these newsrooms from the AIJC development grants was just $5,000 per team, and additional funding for the project was scarce and likely also in the range of a few thousand dollars. Few of the participants had any experience with AI prior to beginning the program, and most teams had to balance the time they spent on their AIJC project with their day-to-day jobs in often frantic newsrooms.

Most of these barriers did not appear to be significant factors in limiting what the AIJC teams were able to achieve. The geographic dispersal of the teams proved essentially irrelevant, as all teams were fully engaged in the global digital community and at ease with accessing and using digital tools and using digital resources for learning and troubleshooting. The smaller size of the newsrooms appeared to be an advantage, rather than a disadvantage, because the teams were able to make quite significant implementation decisions in a largely self-contained way or with minimal bureaucratic overhead. Access to AI generally meant either an API account for OpenAI and a budget for tokens, sometimes even just a few paid accounts for ChatGPT Plus, or sometimes an account for ElevenLabs or MidJourney. Access to some tools was more expensive, for example Rappler's use of tools from a video automation vendor, but even this was in the low thousands of dollars. Some participating teams did have developers or consultants associated with the project, which was more expensive, but even these represented relatively short, well-contained projects with limited investment. None of these teams had resources that were even remotely equivalent to what large legacy news organizations routinely invest in innovation projects.

There was one potential barrier to realizing ambitions for deploying AI in several of the participating teams, which was the need for infrastructure that enabled

optimal use of the outputs of AI. This included a desire for content management systems that could work with AI-generated content, and content serving systems that could handle the personalized serving of different versions of content to the audiences for whom they were intended. But even here we observed creative "workarounds" that enabled progress in spite of that infrastructure, such as Rappler's decision to set up a new brand for content targeted to a new audience.

The most ambitious AI projects may not be the riskiest.

The projects across the AIJC cohort represented different levels of ambition, ranging from some quite safe and incremental initiatives to some surprisingly ambitious projects, and there appeared to be little correlation between level of ambition and the success of the project. While we did see a clear example of a very ambitious project that fell short of a fully implemented application, we also saw several other examples of ambitious projects that moved forward towards useful implementation. Cuestión Pública's Odin project is one such example, as was Rappler's TL;DR. Similarly, we saw at least one example of a relatively incremental project that was slow to develop, and which ultimately failed to deliver meaningful results to the newsroom. These examples—albeit from a very small sample—suggest that setting ambitions to fundamentally, and even dramatically improve some aspect of a newsroom's journalism or impact might be quite reasonable, and might produce motivation, imagination, and organizational engagement that enables big things to happen relatively quickly. Likewise, the over familiarity and relatively low stakes of minor incremental projects might act as disincentivizing influences that reduce the probability of success in applying AI.

Coding is not always needed to automate workflows with AI.

The most complex and ambitious AIJC projects usually had some support from developers or technical consultants, however we also saw examples of very complex workflows implemented through the sophisticated use of no-code tools, or even just with thoughtful use of user interface implementations. One example was Rubryka's project, which used the Make.com no-code workflow automation tool in coordination with the AirTable no-code database platform and carefully constructed API calls to GPT-4 for step-by-step tasks. Another example was

Zamaneh Media's UI-based workflow, including the use of a no-code user interface tool to expose the functionality of their Newsletter Hero application to journalists.

This observation is confirmed by the increasing use of "no-code" development in the wider technology community, with many examples of scalable, professional products being built and deployed using no-code techniques. This trend predated ChatGPT but has been accelerated by the appearance of LLMs that can essentially take care of quite complex operations that would previously have required some Python or SQL coding to implement. The ability to create simple AI agents without code, using OpenAI's "Custom GPT" was not available during the AIJC program period, however this development further simplifies the ways in which relatively sophisticated multi-step workflows can be automated.

Furthermore, even in cases where "no-code" approaches are insufficient, it is increasingly possible to rely on LLMs as coders and "coding coaches" that can produce complex, working code and that can help a person without coding skills to understand, troubleshoot, and implement that code. The experience with "no-code" techniques observed in the AIJC program suggests that these may play an increasingly large role in AI implementations, especially in small newsrooms.

Some AI projects imagine new kinds of journalism.

Most AIJC projects, unsurprisingly, sought to automate tasks that were already part of an existing manual workflow, for example the impact tracking workflow built by Daraj or the SEO workflow built by Raseef22, or sought to use new AI-based workflows to produce news products that are relatively common in journalism, such as Scrolla's text articles, Meduza's summaries or even Rappler's comics and videos. A few projects, however, imagined entirely new ways of approaching the core journalistic mission of the organization—essentially inventing new ways to do journalism.

An excellent example of this is The Initium's project, which sought to enable the automated and systematic monitoring of a large and dynamic corpus of global news sources in the Chinese language, for the purpose of identifying emerging news stories and using the breadth of source materials to identify patterns and trends. Furthermore, The Initium team began the project with an even more imaginative vision for how they might better deliver their journalistic mission within an end-to-end AI-enabled value chain. This vision consisted of the

newsgathering portion that became The Initium's project, combined with an AI-enabled news production portion that reversioned content to make it more accessible to more audiences, similar to Rappler's project. Between the AI news gathering and AI news production portions would be a kind of AI-assisted "editorial switchboard," in which editors would manage, control, and verify the identified stories and resulting communication experiences. Furthermore, this imagined end-to-end system would also include the systematic and partially automatic interpretation of news from sources controlled by the Chinese government in ways that could "read between the lines" to produce useful news from censored sources. The Initium team was only able to attempt a small portion of this vision during the AIJC project, however their thoughtful articulation of this overall vision demonstrated the extent of the potential opportunities available from these tools.

A second example of a particularly visionary project is Cuestión Pública's Odin, which was described in detail earlier. Although Cuestión Pública's team was motivated solely by their own journalistic mission within Columbia, they nonetheless imagined an entirely new way of combining data journalism and breaking news into an "always-on" journalistic lens that could contextualize breaking news with authoritative structured data and then communicate that on social media with a distinctive editorial voice. Their approach is generally applicable to any kind of journalism in which grounding and contextualizing current events in a detailed and specific historical record is valuable. Using the Odin project as a demonstrator, it is quite easy to image that same technique applied at vast scale, applied across many areas of human society, delivering transparency of political, economic, and cultural life that vastly exceeds anything that we currently experience from journalism.

A third example of a new AI-enabled vision for journalism is Scrolla's vision of using AI-based media production tools and AI-centered training and support to expand the range of journalists whose reporting they can publish, empowering people rooted within overlooked and marginalized communities to do original reporting and produce content that they would not otherwise be able to do.

These examples show that the application for AI to journalism need not be restricted to how we presently conceive of news or journalism, or how we presently produce news products, or even to the products that we presently produce. AI has given these newsrooms an opportunity to fundamentally reimagine what service or value they are providing to their audiences and to envision practical ways to deliver that new value.

Prompt editing/engineering is central to most AI projects.

The most common theme across all the AIJC project teams, by far, is the centrality of "prompting"—detailed and specific control of the language model—to successfully applying AI. In all the essentially successful AIJC projects, the participants became skilled "prompt editors" or "prompt engineers," and most became increasingly fluent in the iterative, research-informed and thoughtful development of sometimes quite sophisticated prompts designed for narrowly defined applications. We repeatedly observed teams initially struggling with prompting, then developing their skills and confidence with prompts and finally becoming almost casually confident in their ability to work with the models to get what they wanted from it. It was almost like watching these teams reach a point at which the language models "joined their team." Simultaneously, we also often observed the relative importance of prompting within the projects change over the program period, from being just one component in a multicomponent system to becoming the central function.

By the conclusion of the project, prompting had become so central that most of the projects could essentially be divided into just two broad components: the prompts themselves and the surrounding activity that supported the prompt and its output. Prompts took on an increasing proportion of team attention as the projects progressed, and many teams developed quite sophisticated prompting strategies, such as multiprompt sequences, assembly of prompts, elaborate system prompts, and provision of few-shot examples as their requirements for model output and their confidence in prompting grew.

There really was no shortcut on this journey. Teams did receive some training and coaching in prompting in the early phase of the program, however it was only through sustained hands-on prompting towards specific task-focused objectives that they developed their own sense of how to approach and evaluate prompting. No team, to our knowledge, developed their prompting skills primarily through theoretical or academic training, although ad hoc examples, tips and coaching throughout the program often accelerated their progress. Other than just developing hands-on familiarity and confidence in prompting, the greatest challenge to teams in improving their prompts seemed to be in carefully defining the tasks they wanted the model to complete. Understanding specifically "'what to ask for" appeared to be a non-trivial requirement, even for familiar tasks. Teams did, however, become noticeably better at this over time.

An excellent example of this journey was Zamaneh Media, discussed earlier in this report. The two members of Zamaneh's team began the project with essentially no prompting skills. Following the introduction in the educational phase of the program and in early mentoring, they tentatively began developing prompts to deliver on their chosen task. This appeared to be quite frustrating in the early weeks of the build phase, due to the mismatch between the intentions behind specific prompts and the actual results that the model could deliver as a result of those prompts. The team continued their hands-on engagement, however, and gradually became more and more comfortable in specifying what exactly they needed from the model to complete a task, in formulating prompts that expressed that need, and in evaluating the resulting model output. The prompting and model output also took on a larger and larger role in the mentoring discussions over this time, and, clearly, within the project itself. The interplay between increasing confidence in prompting ability and the increasingly important role of prompting within the project was clear over the project duration.

Another example, this time from a more sophisticated starting point, was Meduza. This team had some initial familiarity with AI and had skilled developers as part of their team, and they began the project with more confidence than many of the other teams. The Meduza project was, however, particularly ambitious involving the detailed summarization of extremely long and very complex and nuanced articles written in Russian into much more readable shorter documents in English, while also retaining the voice and standards that have characterized Meduza's journalism. This was much more than a simple summarization task, and it soon became apparent that getting the essence of those longer articles into the summaries in a controlled way, while retaining some of the nuance and subtlety they contained, was not going to be straightforward. As a result, the project endured a period of frustration several weeks before the completion of the build phase during which they question whether the model was capable of the task. This caused them to engage in an intensive, focused period of prompt work, by the end of which they had achieved much more complex prompts that delivered much of what they needed for their summaries. This team was simultaneously pursuing model fine tuning as an additional approach to improving their ability to summarize their articles. However, it was very clear that the feasibility of the entire project hinged primarily on the team's ability to construct and evaluate quite sophisticated prompts.

Much of these observations of the central role of prompting in the development of the capabilities of the teams and their projects was diffuse, happening gradually over many weeks and expressed in many different ways. There was one particular

"marker," however, that tracked this progress—the decline of the idea that a model "could" or "could not" do a particular task. This idea, expressed in different ways, was very common early in the program, and, of course, there are indeed a great many things that LLMs cannot do and perhaps will never do. But over the months these teams worked on their projects, many of them came to question whether the insufficiency of a prompt was due to the model or due to their communication with the model. It was in some ways like watching a newly promoted manager gradually becoming familiar with managing a new junior employee, and exploring whether unsatisfactory work was due to the employee or dueto the way they had been instructed.

Some advanced AI techniques are within the reach of small teams.

We were surprised by the resourcefulness of some small teams in deploying their resources to take advantage of quite advanced AI techniques, including Retrieval Augmented Generation (RAG) techniques based on document embeddings and fine-tuning of models using proprietary data. At the time of the AIJC program, these techniques required skilled developers, but the fact that they were considered and used by several AIJC teams suggests that their use may be accessible to other similarly resourced newsrooms. This accessibility has improved substantially since the completion of the AIJC program, for example by OpenAI's easy GPT-3.5 fine tuning functionality, or by the appearance of "custom GPTs" that enable RAG-like use of source documents in their construction.

RAG was used by Cuestión Pública, and was designed and implemented by a technical contractor that Cuestión Pública employed for the project. The corpus used for the application was the large, structured dataset of Colombian political and business elites that was at the center of Cuestión Pública's project. This corpus was converted to vector embeddings at difference scales, and then used in a retrieval process in which a vector embedding of a document containing "breaking news" information acted as a query. The system then returned the set of relevant embeddings and underlying structured records which could act as a smaller, focused input to the language model producing the draft output for publishing. This was an ambitious approach, and was not without its technical difficulties. However, the basic infrastructure needed to implement it (such as an embedding API, the Elasticsearch vector database) were readily available.

Another example of a team deploying an advanced technique is Meduza's use of fine tuning to improve the ability of LLMs to produce acceptable summaries of their very long, complex, and nuanced articles. Fine-tuning is essentially a technique for improving a model's performance on a particular task by training it using a relatively small number of examples of inputs and outputs from the task, and is applied as a smaller, secondary model whose influence is essentially combined with the foundation model at inference time. Meduza had software developers on their team and were well-equipped to undertake fine-tuning. Furthermore, they were well aware of the trade-off (at the time of the program) between the capabilities of leading foundation models like GPT-4 and the smaller, less performant models for which fine-tuning was then possible. The Meduza team also had a clear understanding of their task, of why they wanted to apply fine-tuning to that task, of the training data they would need to conduct the fine-tuning, and of the potential for fine-tuning to deliver improved performance that was likely not available through advanced prompting—a necessary understanding that the Meduza team had worked hard to achieve.

Several other teams explored fine-tuning and other advanced techniques during the program, although these were not central to any projects other than Cuestión Pública's Odin project. The fact that we encountered several such efforts, by small and relatively low-resource newsrooms during the AIJC program suggests that these techniques are likely withing the abilities of most newsrooms to implement *if that were to be useful*. As was clearly demonstrated by both the Meduza and Cuestión Pública's examples, advanced techniques require well-understood tasks, a reason for their use that techniques like few-shot prompting cannot address, a useful dataset, and the technical ability to execute the implementation. As we have seen, however, some small newsrooms can meet those requirements.

Facilitation can accelerate AI adoption in newsrooms.

All the AIJC teams already had the necessary access, abilities, and resources needed to begin exploring and applying AI to tasks and opportunities in their newsrooms, probably even without the relatively small AIJC development grant. They also generally had an idea of a problem or need that they wanted to solve or improve using AI. Most of them, however, initially seemed to lack a way to get started—a way to discuss, clarify, and define their problem in terms of AI, to understand AI models as specific tools rather than "wishing wells" and to plot a

path to pragmatically applying the tools to their problem. They seemed to lack a frame for applying AI. This situation often seemed to be merely a matter of confidence—i.e., of achieving enough familiarity and comfort with practical AI to be able to become self-supporting in learning, exploration, and experimentation. The AIJC program offered two forms of facilitation that clearly helped teams to get started and develop confidence—education and mentoring.

The education phase of the AIJC program, and particularly the compressed "boot-camp style" timeframe for it, provided an excellent immersion in the intensity of the AIJC program—an intensity that persisted throughout. The series of presentations, many as part of the JournalismAI Academy for Small Newsrooms, provided a baseline of background awareness for the fundamentals of AI and the opportunities for its application in news, as well as a common vocabulary and a contextualization of applications within surrounding concerns like ethics and safety. It also exposed the teams to accessible people who were already applying AI, thereby normalizing the practice for some participants. All that was vitally important, but perhaps the largest contribution of the education phase to the program was as an "initiation"—a ceremony that inducted participants into practitioners of AI rather than just spectators.

Providing useful and impactful feedback and advice to the teams, usually in the weekly one-to-one meetings or during special meetings to address specific issues, required the mentors to maintain an awareness of each project, and its objectives and its trajectory. Knowing in some detail what each team was trying to do, how they were approaching it, what they had already tried, what they were stuck on, and what their strengths and weaknesses were enabled feedback and advice to be relatively specific and helped maintain momentum during the design and build phases. Mentors were not part of the teams, but they needed to be quite close to the teams and their hands-on activities to be useful.

Incentives are important for fast progress with AI.

The AIJC teams were motivated, skilled people who would likely have made progress in applying AI regardless of their participation in a managed accelerator program. We repeatedly observed, however, that the incentives provided by the AIJC program were instrumental in helping the teams to make rapid and focused progress. We also observed that small amounts of practical, focused feedback or advice could often have a disproportionately large impact on the direction pf projects.

The fact that the program was structured as a competition was an important incentive, from the initial formulation of proposed projects during the application process, to the competition among the twelve participants to become one of the five finalists, to the final competition to win a substantial grant. We often saw and heard team members refer to the competitive situation as a motivator, often at critical moments.

The structure of the program, with weekly check-ins, periodic deadlines, and public sessions where progress was shared also seemed to help teams make steady progress. We worked hard to make sure that we did not distract teams from their work, but there was also no point during the program when teams were left alone without guidance or a sense of what was next. This was a balance, perhaps not always achieved successfully, but in hindsight this structure likely enabled accelerated progress in most teams.

Finally, the sheer pace of the program seemed to help teams retain focus and dedication. The initial "boot camp" of compressed educational sessions, the initial project definition deadline, and the relatively brief development phase before the judging sessions all served to incentivize the teams to maintain momentum and move forward quickly.

We could have done some things better.

The AIJC program was a learning experience not only for the participating teams but also for the organizers and mentors. In general, the overall design of the program seemed to be quite robust and successful, and the usefulness of the accelerator-style approach to building AI capacity in news organizations was clearly validated by the program. Nonetheless, there were multiple areas where we felt the program could have been improved.

The communication between organizers and participants could perhaps have been better. Most communication was informal and ad hoc, which might have left some gaps, given the relatively large number of teams and team members. Some kind of combination of a more managed communication process for "housekeeping" and ad hoc, informal process for team-by-team engagement might have worked better.

The logistics of dispersing the development grants to participating teams could have been faster. Some teams were unable to procure the resources they needed to implement their projects due to delays in grant disbursement. The intent of the funding component of an accelerator program is to pay for the costs of the project,

and delivering funding well before the start of the build phase is necessary to enable this.

We did not sufficiently enable opportunities for the entire AIJC cohort to share experiences, form a community or even network with each other. There were a few attempts to do this, such as a cohort Signal channel and in some cohort-wide video calls, however those were not well used. Forming relationships across teams is challenging to do through predominately online engagement, and the competitive structure of the program sometimes works against collaboration. However, there might be ways to facilitate community building within the existing program structure. One of these might be to move more or most of the communication onto Slack or Teams. Another might be to provide a way for teams to join forces during the program, for example based on application or circumstances, without sacrificing their grants or opportunities to win. Another option might be to provide an in-person initial workshop to either kick off the program or to present project designs following the education phase.

That last possibility might also address another challenge that was sometimes voiced by participants early in the program—that of too much time spent in video calls during the educational phase of the program. The intensity of video calls during this period was significant and this burden might have been reduced by keeping these calls more focused and pragmatic, by moving some to in-person events, or by moving some of their content to fully asynchronous videos or chat conversations. It should also be noted that the "boot camp" intensity of the initial engagement of an accelerator program is a key way in which participants get into the mindset needed for executing their fast-paced projects, but perhaps we could find a better way to do this.

There are some things we do not know yet.

As organizers of the AIJC program we are, of course, biased in our interpretation of how the AIJC program delivered value to its participating newsrooms and to the wider journalism community. As described above, our general assessment is that the program was a success, and that it clearly accelerated the capacity of its participating teams in applying AI in their newsrooms. Whether the participating teams are successful in applying AI in the months and years after the program is uncertain, however, because there is a lot we still do not know.

- *We do not know how much these AI projects will scale*. Although AI sometimes seems to promise unlimited scale, in practice its use is limited by modal limitations, hallucinations, editing requirements, organizational constraints, and technical limitations in publishing its output and other factors. We do not yet know the degree of these limitations and how they will vary by application area and over time, as models improve (or degrade).
- *We do not know how audiences will react to content produced with AI*. By the conclusion of the AIJC most of the participating teams had either tested the output of their prototypes with audiences or published content produced with AI assistance, but the results of these tests or publishing were still unclear due to small sample sizes and the early stage of the new workflows. Clear signals from audience analytics about the value of this content likely will not be available until this content constitutes a larger portion of output, for a more sustained period of time, and is subjected to iterative improvement.
- *We do not know how easily AI workflows will transfer into routine production*. The participating teams were fully vested in the opportunities available from AI, and they all had the support of their senior leadership in implementing these workflows. Yet, some of them may experience cultural, bureaucratic, organizational or technical barriers to fully and permanently deploying their workflows in production. This uncertainty does not apply to those participants whose senior decision-making leadership engaged directly in the program, including Agência Pública, Cuestión Pública, and others.
- *We do not know whether the participating organizations will build on AI as a general capacity*. All the participating teams made clear and substantial progress on their projects, and all clearly developed considerable fluency in prompting, abstraction of tasks, and other general capacities necessary for broad application of AI, but we do not yet know whether those capabilities will diffuse throughout the participating organizations and be applied to other workflows and products.
- *We do not know whether the participating organizations will be able to retain their new AI-capable staff.* We have already seen several participating teams lose key members of their teams, often to technology companies rather than other news organizations.

Recommendations

The AIJC program was intended to provide guidance to stakeholders beyond just the program organizers and participating teams—including: any news organization of any size that is seeking to apply AI in their work; facilitation teams either inside or outside of newsrooms seeking to help or encourage others to apply AI to news work; investors in programs and projects intended to enable or exploit the application of AI to journalism; and to the journalism ecosystem generally, including its scholars and its broader stakeholders in institutions, governments, and societies.

This report is the main way in which that intention is realized, and our recommendations for stakeholders are provided here. These recommendations are subjective interpretations based not only on information about the program described in this document, but also on observations from participating as a mentor in the progress of participants week-to-week.

For journalism providers who want to apply AI:

- *Engage with the tools*. All of the AIJC project teams began their projects in the same way—by sitting down at a computer and logging into an AI tool. There is no substitute to signing up for accounts, learning your way around the user interfaces and APIs, playing with prompts, critically evaluating outputs, and iterating again and again. Applying this new generation of AI to journalism is clearly a hands-on activity, more like riding a bicycle than learning calculus. Developing a visceral understanding of its capabilities and limitations requires engagement.
- *Work on real problems*. There is real value in attempting to use AI and LLMs to solve real problems from your real work—even if you and your organization are just learning to use the tools for the first time. If you focus on issues such as quality, customer experience, editing criteria, and journalistic mission as you learn, then you will be forced to contend with challenges that "toy problems" do not provide. Real-world tasks provide the motivation, the pragmatic constraints, and the nuance needed to eventually deploy AI usefully. We repeatedly observed the AIJC teams encounter difficult challenges as a result of their profession's objectives, and then learn substantial lessons in overcoming those challenges. Users who are learning and deploying AI and LLMs need to work on real problems from the start.

- *Take prompting seriously*. It was clear to us that comfort with developing prompts to meet specific, defined objectives was an essential skill in all of the AIJC teams. Without exception, it wasa skill that took time to develop and that was not always straightforward to achieve. Prompting is about much more than just learning what to type into an LLM, although it is certainly that. It is also about learning how to choose tasks for LLMs to do, about establishing quality criteria for those tasks, about evaluating the output of the LLM against those quality criteria, about managing the resulting feedback, and about learning and improving from that feedback and much more. Prompting is very much like writing—any random literate person can prompt ChatGPT, but any random literate person cannot write valuable journalistic content. That skill can only be developed by hands-on practice, by reading, and by thinking about what you write, and by judging it. The same is true of prompting.
- *Learn to iterate*. No writer or software developer sits down and produces perfect output on one pass. The ability to act, evaluate, adjust, then act again is not only essential to prompting but to deploying AI generally. Not all the AIJC teams were initially comfortable with iteration, and some initially assumed a "waterfall" approach to their projects—write a detailed specification and design document and then rigidly execute that plan. It became clear, however, that teams that either had, or developed, a comfort with constant improvement through the repeated application of feedback made the most rapid progress. It seems that the uncertainty around AI functionality and its application to journalism strongly favors such an exploratory approach to progress towards a defined objective.
- *Take some risks*. It was a risk for the AIJC teams to participate in the program, and to devote as much time as they did to it. Each project began as a risky proposition, with unclear prospects for success. Not all of the projects achieved the objectives that their teams had for them. All of the teams had failures, sometimes even major failures requiring significant readjustment. Nonetheless, every AIJC participant left the program with a substantially broader and more viscerally pragmatic understanding of AI and its application in their newsroom. There is probably no path to exploiting AI in journalism that does not involve risk, and so developing a comfort with risk and expertise at managing risk is probably unavoidable.
- *Be wary of limiting yourself*. In our experience observing the AIJC project teams work through their applications, it was repeatedly clear that the most common barrier to achieving a successful outcome from using AI was lack of confidence, and especially lack of confidence about technical ability. A

significant part of the early portion of the program therefore involved coaching participants to become familiar not only with the tools but with the new power that those tools provided. Again and again, we observed a "lightbulb moment" when a team realized what they could do and quickly expanded their idea of the applications and potentials of AI in their work. An organizational objective for journalism providers should be to get their employees to that "lightbulb moment," where their confidence and their self-sufficiency with applying AI increases rapidly.

- *Reach out to others*. Almost all the achievements of the AIJC project teams and program came through interactions between people. Almost all the disappointments came through isolation. This was mostly true within teams, but also true within the participating organizations and within the industry more broadly. More narrowly, this was also the case with the relationships provided by the program, from individual and team mentoring to the contacts provided in the educational phase of the programme to accessing the broader networks of the Open Society Foundations, the program team and even the competition judges. The application of AI in journalism is generally not a private activity, filled with proprietary information, trade secrets and non-disclosure agreements. Instead, it seems to be a collective activity, with individuals, teams, and organizations eager to learn and to share what they have learned. As a result, the teams and individuals that reached out and connected the most were generally the teams and individuals that made the most progress.
- *Don't give up!* Most of the AIJC teams went through periods of deep frustration when the way forward was not clear or when nothing seemed to be working. In hindsight, these moments were often inflection points when teams were forced to reckon with key aspects of their approach that were not working—sometimes technical, or product-related but sometimes more subtle or even philosophical. It was in working through these moments that leaps in progress were often made, and real confidence was developed. Learning to apply AI pragmatically to news clearly is not easy, but with perseverance and honest appraisal it is clearly within reach of any newsroom.

For managers of AI accelerator programs:

- *Have a detailed written plan for your program*. Design your program carefully, with clear written objectives and a structure that meets those objectives. The written plan for the AIJC was just three pages long but was a critical part of both designing and executing the program. Developing the plan forces you to think through the step-by-step reality of each part of your program, and to develop agreement about it with stakeholders on specifics rather than in terms of vague concepts. You might (and likely will) decide to deviate later, but if you have a clear plan, you will be able to do so thoughtfully, from a firm foundation, rather than drifting aimlessly. Developing the plan also forces you to be very specific about the outcomes you are trying to achieve from the program, which will become the guiding "north star" throughout its execution (for AIJC this was "pragmatism and transformation"). There are many excellent sources of learnings and inspiration for designing AI accelerator programs for journalism, especially from the start-up accelerator community. Y-Combinator, for example, is both particularly successful in achieving meaningful outcomes and also particularly active in sharing its approaches. At a practical level, the foundation of a program is its detailed schedule. For AIJC, we approached this as steps towards an objective rather than as programming an experience.
- *Select project teams carefully*. The primary criteria of team selection should be Whether or not the team can succeed in usefully implementing AI in their newsroom. It is difficult to assess these criteria from a relatively short text application, however evidence of thoughtfulness, effort, earnestness, and pragmatism are usually apparent in an application. It is, of course, critical to consider factors such as the language, situation, organization size, and relative disadvantage in selecting participants. Yet, those constraints are usually readily apparent or can be explicitly requested. One way to ensure consistent selection is to involve multiple people in the selection process, using clear and specific criteria. The selection of participants for the AIJC program was done by a small group of diverse evaluators independently scoring applications and then collectively discussing "edge cases" and erring on the side of diversity of applications—all while keeping the primary criteria firmly in the forefront. Judging by the performance of the participants, it seems to have worked quite well.
- *Set your teams up for success*. Teams participating in a program like AIJC have a lot to consider and to process as they engage with AI tools and try to

pragmatically fit AI functionally to real tasks in their newsrooms. It is important, therefore, to provide for as many of the basic needs of these participants as possible, thereby "freeing up" their scarce attention to engage with the project. In practice, this means providing project teams with enough funding to cover the costs of participation, sufficient training, and education to feel well-prepared for their project, and enough access to expertise and advice to work through basic issues, especially during critical periods such as project design or troubleshooting. It also means providing clarity and specificity about the program and its expectations for the participating teams. Reducing the friction of participating in an accelerator will increase the focus and thoughtfulness that the participants can bring to their projects.

- *Keep a fast pace*. Pace is the secret ingredient of successful accelerator programs, from military "boot-camps" to Y-Combinator. A constant drumbeat of touchpoints, from start to finish, helps to build and maintain momentum, retain focus, and stretch both imagination and implementation efforts. In the AIJC program these touchpoints included a series of presentations and workshops during the educational phase, the all-teams presentation of project designs, the weekly engagement with mentors during the build phase, the finalist judging event, and the final judging session in November 2023. Throughout the program, the teams were generally never without a "deadline" of some kind in their immediate future. This was important because the teams, many of whom had "day jobs" in their newsroom, were generally never able to put aside their AIJC project or shift their focus to other matters for very long. Instead, they were constantly engaged, from start to finish, keeping their attention on the concepts, approaches, challenges, products, and outcomes from their AI projects. This requirement for constant focus is particularly important for developing familiarity and competency in an entirely new fundamental capability such as AI, because it requires knowledge, assumptions, attitudes, and mindsets that are different from traditional journalism.
- *Focus on well-defined projects*. The most critical phase of the AIJC program was the two-week period between the end of the educational sessions and the all-team presentation of project designs. It was during this period that most of the projects took shape, although some teams came into it with well-developed designs and some continued tweaking their designs well into the implementation phase. The key factor in this design period was the increased specificity of designs—the general transition from often vague ambitions to specific descriptions of outcomes, products, workflows, and approaches. Very

loosely speaking, this period marked the transition from thinking and discussing to actually building, driven by a requirement for specifics. Vaguely defined projects cannot be implemented, and therefore an explicit transition to specificity is unavoidable.

- *Focus on identifying and removing blockers*. Accelerator programs should, as described above, set their teams up for success from the beginning. This ethos should also apply, however, to the week-to-week support provided to the teams from mentors, expressed through actively identifying and helping to remove blockers. The identification of impediments to genuine progress should be the central function of mentoring, closely followed by working with the participants to remove or get past those impediments. Sometimes impediments are small and tactical—an approach needs adjusting, or a prompting technique needs to be explored.—But sometimes they are large and strategic—an approach needs to be reconsidered or an ambition curtailed. In many ways, identifying and removing is the central service of an accelerator program to its participants, and is largely what separates such programs from others that offer merely funding, education or convening.
- *Strive for both practicality and transformational ambition*. The simultaneous pursuit of projects that exhibited both pragmatic solutions to specific newsroom challenges and also transformational potential commensurate with the newly available functionality from AI and LLMs was the central objective of the AIJC program. These seem contradictory, because practicality is often associated with incrementalism and transformational innovation is often seen as merely conceptual and aspirational. Nonetheless, it was the explicit pursuit of both of these characteristics in selecting participants and in week-to-week execution that helped the program generally avoid "small potatoes" projects that had no prospect of impact even if successful, and one-off "pie-in-the-sky" prototypes that had no prospect of actually being used in routine, day-to-day publishing in newsrooms. By explicitly including both criteria in our objective, we essentially biased the AIJC program towards actual real-world impact from the application of AI.
- *Provide accountability as a service*. One small but important value provided to participants in an accelerator program like AIJC is accountability—small ways to help participants meet the goals they have set for themselves by participating. This includes soft and hard deadlines, metrics or other forms of evaluation against goals, clear articulation of expectations, and open and frank discussion of issues. Cumulatively, these small expressions of accountability help to keep participants honest with themselves, much in the same way as a

weighing scale helps dieters keep to their meal plans. Applied gently and with understanding, they combine to help participants make the most of the opportunity offered by the program.

- *Execute the program logistics well*. This recommendation is obvious but is worth explicitly stating, nonetheless. A program like the AIJC is made up of lots of small tangible actions. The more smoothly those actions occur, the more attention everyone involved in the program can place on the core tasks and objectives. Setting up meetings, planning documents, educational sessions, mentoring sessions, grant disbursement, travel arrangements, communications, are all critical. Arranging all this is perhaps not the most exciting aspect of an AI accelerator program, but the individuals who do these tasks should be appreciated and assisted.
- *Be unlimitedly accessible*. It is not always easy for program participants to know when or how they will use the mentoring or logistical services offered by an accelerator program. The AIJC program attempted to offer its participants access to expertise whenever they might need it, regardless of their time zone, schedule or amount of advance notice. Engagement with mentors outside of scheduled sessions was rare but was important when it occurred. Furthermore, several participants mentioned that just the knowledge of the availability of support or advice if they were needed was important to their confidence as they explored functionality and engaged with problems.
- *Share, genuinely, as much as possible*. Throughout the AIJC program we attempted, not always successfully, to share as much information about the program and its contents as possible with the participants. Some of this was via formal communications but much of it was informally, during one-to-one conversations, team mentoring sessions, and small group meetings. It was in these less formal settings that openness, earnestness, and authenticity had the most significant impact, not only on building understanding but also on building trust and safety during what was sometimes a stressful time. These benefits were not only in one direction, but were also vital in enabling the program organizers and mentors to understand and appreciate the reality of the participants' situation, enabling them to adjust and adapt to best serve the participants' needs. Formal communications have their uses, but open, empathetic conversation was the communication style that enabled progress.
- *Meet in person for critical moments, if possible*. The AIJC program was entirely online, except for in-person meetings among the five finalists at Spice Beta in Chiang Mai. Although we did attempt to foster conversation between teams during the program (via Signal and all-teams video calls, for example),

these were generally not successful. Online participation is unavoidable for a global, multimonth program like the AIJC, and the costs of convening in-person are not insignificant, however, in hindsight we probably should have organized an in-person session early in the program. Such a meeting might have enabled relationships to develop between participants, resulting in deeper sharing between teams and a more valuable network that might continue after the program. The most logical point at which to convene teams was at the project definition presentation following the design phase, when most teams have developed a deeper understanding of their projects and all of the teams have a shared educational basis to build on. In-person relationships are a key value provided by start-up accelerators, and the same would likely apply to accelerators focused on AI in journalism.

For investors seeking to support the application of AI to journalism:

- *Accelerators work well for AI in journalism*. Applying AI to journalism requires a new set of assumptions, perspectives, and attitudes about how journalistic work can be done. It therefore benefits from a fast-paced semi-immersive experience in which the friction of engaging is reduced, and support is close at hand. The accelerator concept was originated to provide commercial start-ups with exactly those conditions and has been very successful in moving start-up teams from being interested to being fully engaged and productive within a relatively short period. We saw these benefits of the accelerator approach play out over the course of the program, and it is clear that it can work for mission-driven organizations as much as for commercial companies.
- *Add value*. The Open Society Foundations and its staff were a critical contributor to the value of the AIJC and went beyond merely setting it up, providing funding, and helping manage the program. The Foundations' considerable network was vital in recruiting motivated and capable participants, and in setting up and operating the evaluation and judging processes. The experience of Open Society staff in evaluating journalism investment projects also contributed considerably throughout the program, through feedback and suggestions for particular teams and projects. Finally, the the Open Society Foundations also provided a high-level perspective for the AIJC program, situating it within the landscape of other efforts to help

journalism adapt to AI that was distinct from these other efforts and that also directly influenced the week-to-week operations.

- *Prepare an investment pipeline*. An additional way in which investors can add value to a program like the AIJC to anticipate the development of projects that will require further investment. We did not do this for the AIJC, but the need for it became apparent later in the program. An example is the Odin project developed by Cuestión Pública, which had the potential to significantly expand the watchdog function of public service journalism. While the project received an honorable mention from the finalist judges for its ambition and significance, there was no easy way to seamlessly move it forward with additional funding. A pre-established funding pipeline that anticipates such situations, perhaps across multiple investors, would help to maintain the momentum of such high-potential projects.
- *Require accountability*. Throughout the AIJC program we attempted to maintain a relatively high degree of accountability, including during initial assessments, using various deadlines, and via a focus on the metrics by which the impact of projects could be measured—especially on the audiences of participating newsrooms. This had varying degrees of success, but the explicit focus on accountability did raise the overall quality of projects and improved the learning experience of participants. An explicit requirement for accountability is built into commercial start-up accelerators, but it is equally important for mission-driven programs because it brings focus, realism, and honesty to these programs and their projects. Those qualities are particularly important in exploring the application of AI to journalism.

For the journalism industry generally:

Prepare for rapid adoption of AI in journalism. The AIJC program clearly and repeatedly demonstrated that teams and newsrooms can go from a standing start to deep understanding and deployment of AI within a few months. Furthermore, the program repeatedly demonstrated that a lack of coding skills or technical training is not an impediment to developing and deploying solutions based on AI. This apparent lack of friction, combined with the enthusiasm and optimism we observed in all the participating AIJC teams, suggests that newsrooms can, and probably will, deploy AI-based solutions quickly and widely—perhaps within months or a few years, rather than a decade.

Reset expectations about size and resourcing. The empowerment of small newsrooms that we observed during the AIJC program suggests that assumptions

about AI-driven innovation being disproportionally restricted to large, well-resourced newsrooms in the "Global North" might be outdated. We should perhaps instead expect imagination, motivation, and agility to become significant advantages, and assume that some small organizations might move very quickly. Observing the speed of decision-making and the focused conversations on risk tolerance were particularly illuminating during the program.

Expand your conception of journalism. Expect imaginative applications relatively soon as new ways of doing journalism using AI are invented. During the AIJC, we observed this in the ambitious projects pursued by The Initium and Cuestión Pública,, which sought to dramatically expand the scale and speed of monitoring and contextualizing news events. We observed it also in the imaginative workflow designed in the Rubryka project, and in aspects of several other projects. The fact that we saw so much imagination and ambition in such a small sample of newsrooms suggests that considerable experimentation about what journalism can become in an AI-mediated context will occur.

Keep mission and values constant. Throughout the entire AIJC program we did not observe any compromise of a newsroom's journalistic mission or values, and repeatedly observed enthusiasm for furthering journalistic missions and values using AI tools. Unlike the social media era when temptations to compromise in pursuit of traffic were everywhere, and unlike the deployment of AI-first content-scraping aggregation, this does not seem to be the case with the use of AI in existing newsrooms.

Next steps

AI has arrived and will probably change journalism fundamentally. Programs like the AIJC, and the 12 remarkable projects that it facilitated in 2023, are useful for exploring AI's potential in journalism and empowering and inspiring other news producers to apply these tools in their own way to serve their own missions. AI is currently in its infancy, and its application to news is largely unexplored, and the AIJC program was in no way definitive or complete. We hope to learn from it, share it, and use it to try to improve the outcomes of AI for news.

Preparing journalism for AI

The AIJC program not only sought to better prepare its participating newsrooms to take advantage of opportunities from AI, but also to contribute to the preparation of journalism as a societal function for an AI future. This report is one way that we hope to make this contribution.

Developing a better assessment of what worked and what did not in the individual AIJC projects requires a longer-term assessment. We expect to "check in" with the 12 teams periodically to assess their progress, their impact within their organizations, and the new value they produce for their audiences. The new lessons from these check-ins will be published via popular news innovation blogs and social media.

Another way that the experience of the AIJC might help journalism prepare for an AI future is via its tangible demonstration of the opportunity and practicality of applying AI within newsrooms. By using the AIJC projects as real-world case studies, communicating about the experiences of the AIJC project teams and incorporating the lessons from the AIJC into training, strategy development, investment, and industry discussions, we hope to better ground these activities in real-world experiences. Furthermore, the 12 AIJC projects can serve as a small "survey" of attempts to apply AI in news, and their relatively high success rate may encourage other news providers to engage with AI tools or at least to anticipate AI as a significant factor in their operations.

By sharing the program-level lessons with others producing accelerators, collaborations ("collabs") or development programs around AI in news, we hope to increase the leverage of the AIJC with journalism-supporting organizations and investors. AI will be a permanent part of our information ecosystem.The sooner and the more extensively journalism-supporting organizations engage directly with it the better. News organizations will need more programs education,

mentoring, experimentation, acceleration, and implementation of AI within their operations if they are to adapt to the emerging information ecosystem.

Finally, we hope to operate future versions of the AIJC program, including AIJC 2024. The objective would be to not merely re-run the AIJC 2023 program, but to update it based on the experiences from 2023 and on the developments, new functionality, and new understanding about AI that emerge during 2024. Change in the news industry is accelerating.Learning, flexibility, and adaptability are critical attributes for navigating that change. This applies as much to accelerator programs like the AIJC as it does to news publishers themselves. We do not yet have a fixed set of "best practices" for AI in news on which to base a similar accelerator year after year. Continual reinvention is essential.

Anticipating the AI-mediated information ecosystem

The 12 projects in the AIJC program generally sought to apply AI to tasks or products that were already familiar to the participants, which is a useful and natural place to start. But even during the program, many of the participants were looking ahead to a time when AI might be ubiquitous in our news ecosystem. The implications of new AI-enabled influences on the information environment, including generative search, fluent conversational experiences, and widespread access to media-creating AI tools were frequently discussed. Similarly, the long-term potential of techniques being deployed in some of the AIJC projects, such as AI-assisted newsgathering, AI agents executing complex tasks, and proliferating new AI-based consumption experiences were also often discussed.

The future of AI in journalism seemed to be anticipated, roughly, as two phases: an early phase focused on the automation of existing newsroom tasks, producing familiar news products; and a later phase that might be fundamentally different than the current news environment and also much more uncertain. Although the AIJC program did produce some projects that perhaps advance toward that new AI-mediated information ecosystem—Cuestión Pública , The Initium, and Rybruka come to mind—the program was primarily oriented towards pragmatic and immediate application of AI in the participating newsrooms as they currently operate. A different kind of initiative would be necessary to explore the possible ways that AI might change news over the long term.

The AI in Journalism Futures (AIJF) scenario planning program is a follow-on project intended to do just that. Launched in February 2024, this program

combines two different methodologies: a broadly targeted open competition seeking short descriptions of the possible ways in which AI might fundamentally change the information ecosystem; and a formal scenario planning process in which the participants are selected from the winners of that competition.

The objective of the AIJF program is to produce a small set of robust, authoritative, and detailed scenarios that cover the space of possible futures for news under the influence of AI. The intent is that these scenarios will be useful for long-horizon investment planning by stakeholders in journalism, including news producers. Another result of these scenarios is that they will help facilitate an industry-wide conversation about how to prepare for the structural effects of ubiquitous AI on journalism.

The AIJF program will conclude with a scenario planning workshop in mid-April of 2024 and the full results—including anonymized contributions—will be published shortly after. It is likely that future versions of the AIJC program will be substantially influenced by the scenarios produced by the AIJF project.

Some concluding thoughts

This report describes the AIJC program and its projects, outcomes, and lessons, but it is harder to describe the feeling of engaging with these teams at this special moment in history when the promise of AI is fresh and the opportunities for its use in journalism seem limitless.

Throughout this experience, I was continually and deeply impressed by the ingenuity and resourcefulness of everyone I worked with. It is clear to me that journalism is very well-supplied with capable people who can figure out how to use AI, with all of its limitations and risks, to inform publics and hold the powerful to account. History shows that transformations of the magnitude of AI's likely impact on news are unavoidably very difficult to navigate—the printing press, the internet, social media—but journalism clearly has the raw talent necessary to adapt to an AI future. In my opinion, the largest lesson from the AIJC for leaders of news organizations is to empower your people, listen to them, remove the friction they experience and then watch them do amazing things.

Another feeling that I took from my AIJC experience was that well-earned confidence can unlock remarkable things. The confidence that the AIJC teams developed over the program was inspiring, as was their willingness to face their challenges squarely, engage with the hard questions, do the necessary work, and persevere when nothing seemed to be working. These teams really demonstrated that there are no short-cuts or "special secrets" for applying AI, just learning, thinking, building, assessing, fixing, iterating and then, eventually, succeeding. That is probably a good lesson for journalism generally.

Acknowledgements

AIJC is an initiative led by **Shuwei Fang**, Associate Director, Programs, at the Open Society Foundations.

The program was a global collaboration that brought people together to learn how to apply AI towards informing society with truthful, impactful journalism. Each of the people listed here were integral to the operation and success of the program.

Mary Fitzgerald, Director, Programs, Open Society Foundations

Suzanne Clark, Associate Program Manager, Expression, Open Society Foundations

Francesca Edgerton, Special Assistant, Expression, Open Society Foundations

JJ Robinson, Senior Program Officer, Expression, Open Society Foundations

Susan Valentine, Division Director, Expression, Open Society Foundations

Michael Moss, Program Manager, Expression, Open Society Foundations

Premesh Chandran, Regional Director, Open Society–Asia Pacific, Open Society Foundations

Dr Bahareh Heravi, Institute for People-Centered AI, University of Surrey

Rishad Patel, Co-Founder & Head of Product, Splice Media

Alan Soon, Co-Founder & CEO, Splice Media

Gina Chua, Executive Editor, Semafor

Abhijit Das, Program Director, Stichting Democratie en Media foundation

Valer Kot, Senior Media Advisor, Media Development Investment Fund.

Marina Walker Guevara, Executive Editor,Pulitzer Center

Sabrina Argoub, Program Manager, JournalismAI,London School of Economics and Political Science

Tshepo Tshabalala, Manager of JournalismAI,London School of Economics and Political Science

Joanne Kuai, PhD candidate, Karlstad University

Gary Rogers, Senior Newsroom Strategy Consultant,Fathm

Jonathan Soma, Knight Chair in Data Journalism,Columbia University

Uli Köppen, Head of the AI + Automation Lab,Bayerischer Rundfunk

Tess Jeffers, Director of Data Science,*The Wall Street Journal*

Alyssa Zeisler, VP of Product Management, Dow Jones

Nick Diakopoulos, Professor in Communication Studies and Computer Science,Northwestern University

And a special thank you to all the individual members of the 12 teams who worked so hard and so enthusiastically over the months of the AIJC program.

These remarkable people are leading journalism into its emerging AI era.

Author biography

David Caswell was the lead consultant for the AIJC program.. He is the founder of StoryFlow Ltd, an innovation consultancy focused on the application of AI to journalism. He was the executive product manager of BBC News Labs, and previously held senior roles leading AI and machine learning initiatives at Tribune Publishing, the *Los Angeles Times* and Yahoo! Caswell also publishes peer-reviewed research on computational, structured, and automated forms of journalism, including previously as a Fellow at the Missouri School of Journalism. He is a frequent speaker and writer about both strategic and applied opportunities and challenges for reinventing news for the emerging AI-mediated digital media ecosystem.