

ISOJ 2022: Day 1

News and AI (#ISOJ Journal peer-reviewed research panel)

Chair: [Amy Schmitz Weiss](#), ISOJ research chair, **San Diego State University**

- Co-Chair/Discussant: [Seth Lewis](#), professor, director of Journalism Program, **University of Oregon**
 - **Patrick Howe, Christine Robertson, Lindsay Grace, and Foad Khosmood**, California Polytechnic State University, San Luis Obispo & University of Miami: *Exploring Reporter-Desired Features for an AI-Generated Legislative News Tip Sheet*
 - [Daniel Trielli](#) and **Nicholas Diakopoulos**, Northwestern University: *Algorithmic Agenda Setting: The Shape of Search Media During the 2020 US Election*
 - [Colin Porlezza](#) and [Giulia Ferri](#), Institute of Media and Journalism IMeG, University Della Svizzera Italiana, Lugano, Switzerland: *The Missing Piece: Ethics and the Ontological Boundaries of Automated Journalism*
-

Amy Schmitz Weiss Hi. Good afternoon, everyone. Welcome to our research panel on AI and news. We're really happy to have you with us. Today's research panel is focused on the papers from our paper competition we have for ISOJ every year, a blind peer-reviewed competition. And the papers that are here were the ones that were selected from the competition. Our research that is featured in this panel today is also for your viewing online, so you can go to [ISOJ.org/research](https://www.iso-journal.org/research) to check them out for free. Fastest publication ever in terms of conference to publication, so you can check them out there and read all the awesome work. And this year's ISOJ Journal has been guest edited by our UT alum, Dr. Seth Lewis, and he's also director of journalism and program at the University of Oregon. We're really happy to have him guest edit this important journal on AI and news, an area that he's very familiar with and well versed in. And before we start the panel, of course with this ISOJ paper research competition for years, we've also given out a top paper award based on the reviewers scores from the competition. And so we would like to have our panelists come to the stage for receiving the top scores from their reviewers. We'd like to congratulate Daniel and Nicholas for the top paper. I'll go ahead and turn it over to Seth.

Seth Lewis Okay. Thank you, Amy. And thanks, everybody, for being here. We're looking forward to a really great session talking about these really good research articles on AI and news. Just a very brief, personal note, I have really fond memories of ISOJ. I came here for my first one as a first year Ph.D. student in 2007, and I've been to many since then, and Rosental and this conference have been influential shaping the growth and trajectory of my work. And I really appreciate that, so thanks very much. All right. We're going to begin with our first presentation, as you can see there on the schedule, is entitled Exploring Reporter-Designed Features for an AI-Generated Legislative News Tip Sheet, and Lindsay Grace from the University of Miami will be presenting on behalf of the other colleagues on that team. So Lindsay.

Lindsay Grace Thank you. So I have a lot of slides and not a lot of time, so I'm going to talk quickly. But hopefully it'll be useful to you. So basically, we see this as to contemporary issues in news. The first is an issue around a shift from discovery to sifting

through an overabundance of information. And the second is that there's not enough time, as part of a first, to sift through many data sources available. So what's basically happened is it's very easy to produce information, but not very easy to actually view it. So the idea here is that the second issue is competition from technology, so it's a very competitive space. And so we see advertising revenues decay, which then affects our ability to report on state legislature, so the resources simply aren't there. And ultimately, these issues are a problem produced largely by technology. And so what we'd like to do is sort of offer a silver lining. And that silver lining, as Cohen and others had said almost a decade ago, is basically for a public interest journalism to thrive, computer scientists and journalists must work together, and with each learning elements of the other's trade. And so Foad Khosmood at Cal Poly and myself basically led a group including Christine Robertson and Patrick Howe, who is the primary author for this particular paper, we basically created a tool with the larger team across the university, Cal Poly, University of Miami and our Austrian partners.

And what we did was we basically implemented and evaluated a prototype that's an AI-informed news tip sheet, and this basically scrapes public data from California and Florida proceedings, state legislative proceedings, and then looks at screen recordings, looks at meeting minutes, looks at websites, and aggregates that content into a basic sifting system. So it does the sifting of absorbing anomaly and newsworthy events for us. Basically we discover news events, we put them in a reporting archive, and then we create article drafts and these tip sheets. So the system looks like this, and these are the cards that are basically anomalous events, things that are interesting or novel as the system basically identifies them. And then if you drill into each of those, what you'll see is this version of the tip sheet, and the tip sheet really just does a few things. It provides a basic overview giving you things like the basics about the bill that just passed, who's involved, where it happened, observations on the statistical ends that are easily scraped from websites and proceedings, as well as doing some simple math for you, doing some aggregation, and identifying specific elements, and gives you some metadata, where we're correlating across lots of different websites and other public resources to give you drill-down information and a wider perspective. So the idea is that you can get this wide perspective, but then drill down into the specifics that a news reporter or analysts would like.

And the idea in this particular paper was that we were evaluating prototypes, and so what we did was we outreached to 10,000 professionals in the industry. We got 193 responses, and that was generally split out to 45% of them were editors, 36% were reporters, and 8% were executives. And one of the core things we'd like people to recognize is that while 98% of the respondents says it's important to cover the policies and politics unfolding in state legislatures, only 37% felt as though they were well resourced to do so. And so we're trying to fill that gap. We basically got some analysis from them and asked them what their highest priorities were, and so of all the features we offer in this tip sheet, number one was listed here. Number two is sifting and sorting. And number three is data collection, the data dashboard. And we also did 60-minute virtual focus groups with these individuals where we just did deep dive on what they like and what they didn't like. And in this case, we got a bunch of interesting observations. So, for example, one of our participants basically said this would be an efficient way to do better to serve our communities like we claim we want to do with more than just clickbait. These are stories that change lives, and if we're not reporting them, who will? And that was a metro area news reporter. The other thing to note is there are, of course, critiques of this, and there are concerns. And the primary concern for people that we talked to with some version of transparency. They identified that they wanted to know how the algorithm comes to use words like "controversial." They wanted

some sense of the algorithm. They also wanted to know who to attribute summary dashboard information to, so who gets credit for doing this work? Some other observations that were made that we found quite useful. One person mentioned that they only have 90 seconds so it needs to make it in, and it could help if we could do the same job through this product. So there's a question about interest. The other is that there is generally a sense that it could expand how audiences are served, so brainstorming new digital products and new ways to make connections for audiences was another observation that one of our focus group participants made.

And ultimately you can read the entire paper and its details at the ISOJ Journal, which is now available online. And if you have any questions, that's why I'm here. So I'll tell you more about AI for reporters. Thank you.

Seth Lewis Excellent. So timely. Thank you, Lindsay. Wonderful. Our next paper, and I should mention, by the way, that if you do have questions, comments, please post those tweets. Use the ISOJ hashtag to make sure you can be included in the Q&A to come. Our next paper is Algorithmic Agenda Setting: The Shape of Search Media During the 2020 U.S. election. This was the paper that just won the top paper award. Daniel Trielli and Nicholas Diakopoulos from Northwestern University, and Daniel will be presenting.

Daniel Trielli Thank you so much. Thank you for the award. No pressure now to present this at all. Thank you, everyone. It's very nice to be here in my first of this conference that is, you know, organized by a compatriot and also attended by compatriots. My name is in Daniel Trielli, a Ph.D. candidate at Northwestern University. I'm part of the Computational Journalism Lab, and we do these type of study about news distribution and news production using AI. What I'm going to present today is part of this work that we do with the principal in our lab, Nicholas Diakopoulos. Now, it seems surreal for me to be saying this, but I started my journalism path 20 years ago, shortly after the first ISOJ. Back then, that's when I started in my journalism school back in Brazil. In journalism school, we learned a lot of things. We learned about journalistic practice, journalistic values, and our journalistic role in society. And one of the things that I learned back then was the role that we have in helping the public determine the public discussion of the important things. This is formalized in a theory called the "agenda setting theory," which basically states that mass media influences whatever voters think is important. And this is an idea that, if news media publishes the same topics over and over, the voters take that cue to say, "Well, this is an important topic. If news media covers inflation all the time, well, maybe I should think about inflation." Now this is an idea that was initially developed in 1972, and of course, it has evolved a lot throughout the years. Very focused and smart researchers have updated it to incorporate second-level agenda settings, how attributes of the topics, and how they're transferred through that mechanism. The third level of agenda setting, which is the news media, not only transfers the salience of topics, but also the relationships between those topics. But even with all those evolutions, this was a theory that was initially developed for the 20th century journalism, the one that I was trained to do, which is a journalism that expects the news product that we create to be delivered directly to the readers in society, even physically to their doorsteps. But now we know that that's not the case. We have these digital curators and intermediaries between news and the public, and a lot of them are composed by some algorithmic or AI processes. You know, search and social media.

And in our research, we make the case that search is significant, important of those curators that we have nowadays, and we think that because that's what Americans say is important. According to Pew, 63% of Americans say they use search at least sometimes to get their news. And for social media, that number is 48%. So you can see search is this

big intermediary between the news and the reader, and that's one way of looking at this model. Another way of looking at it is something like this, in which we can see search results as its own type of media that is co-constructed between the search platform, the reader, and the news media. So the news media provides the content that is curated by the search engine, and that is read by the reader, and that search is actually initiated by the reader looking for the things that they're interested in. So if this is a new way of receiving journalism, if social media is co-constructed, who sets the agenda then? How much power does news media relinquish in that process, and how much power does the user and the search engine actually have to investigate that?

We came up with these two similar interconnected questions. First, is to what extent did the distribution of topics selected by this new search media just replicates the agenda of the news media? And second, to what extent does the search or input actually alter this distribution? Now you can see all the methodology in the paper that was just published today. And I'm glad to answer questions about it, but basically what we did was we compared news coverage about the political candidates during the 2020 elections in the United States and the search results about those candidates. We contrasted the topics that were extracted by those news media reports and those searches about that candidate, and we compared those topics to see if there's a correlation between the both of them. And for the second question about the power of the user, we basically searched the candidate again, we took the search results, and we did kind of an adversarial exercise with the search engine, in which we search for the candidate plus an interesting topic. Like, for instance, "Joe Biden" versus "Joe Biden environment." Is that going to yield more environment related things or not? And we did the same thing and extracted the themes from the results, and put them into topics and compared them. So here are the results. Again in the paper, you're going to see all the fancy statistical tests we did with the results. But basically, this is a high level display of what we see. And what we see here in yellow/gold, you see the distribution of topics in the news. And then this blue is the distribution of topics in search. And you see that in the rankings, they are kind of similar, except on the news there's a lot more news stories about the elections themselves, the horse race, the campaign stops. But other than that, they're kind of similar in the ranking of importance. You see government things like appointments, impeachments and stuff like that as the top priority. Then we see the economy, race, and all the topics that folks might be interested in.

One thing that is visible is the difference here between the weights, the relative weights, of these topics for each type of media, search media and the news media. Topics like race that are not very represented in news media are very much represented in basic searches about these candidates in the search results. Then for the second question, we compared these candidate names with topics that we might be interested in. And in the first column here, you see that distribution of the basic search about Joe Biden. And if we were to say to "Joe Biden, poverty, hunger, homelessness," one would expect that there would be more information about the economy popping up. And what we see is it definitely does increase, but it's not the main topic still. So we can see a bigger representation of the economy when you search about Joe Biden, and specified that we're looking for an economy-related topic. But still some other subjects, some other topics, are still overrepresented. So this is a topic that we measured, which is specifically immigration, and we actually searched "Joe Biden immigration." Now, you would expect that during the election that's going to be like 80 or 90% of the search results, right? It's actually 5.8%. It is an increase over the zero search results in a basic Joe Biden search, but it's still relatively little. And environment the same thing. There is an increase over the basic environment searches, but still, it's even under things like the economy, for instance.

So the bottom line here is that in that triangulation of who sets the agenda, neither the news media or the user alone have the power to shape the salience of topics in search results or search media. Search media actually take some cues from that, but it also resists that in some ways. That means that search media can drive, shape, or even counteract choices that are made by journalists and readers/users/searchers in the things that they're interested in. So what we do in this paper is we propose a more in-depth exploration of who sets the agenda and who has the power to define what we talk about when we talk about political topics. That was my research. I'm open for questions later, easy questions, during the Q&A that make me look good. And for hard questions, I'll be at the reception later.

Seth Lewis Excellent. Thank you, Daniel. Congratulations again. Our final paper is entitled The Missing Piece: Ethics and the Ontological Boundaries of Automated Journalism, and it's by Colin Porlezza and Giulia Ferri. Unfortunately, there in Switzerland and could not be here to attend. We have video of their presentation, however. My understanding is that there may be parts of the video where Colin may be referring to slides, so those slides are going to be made available on Twitter, I believe, using the hashtag, as well as virtually for those of you on Zoom, you'll get that in the chat. So if I can send it now to the video of Colin's presentation.

Colin Porlezza Hi, everyone. My name is Colin.

Giulia Ferri My name is Giulia.

Colin Porlezza And welcome to our presentation that is entitled The Missing Piece: Ethics and the Ontological Boundaries of Automated Journalism. Now, we actually wanted to show you at the start of our presentation, a special movie from 1957 called the "Desk Set," starring Spencer Tracy and Katharine Hepburn. The movie actually anticipates organizational and professional implications of the electronic computer in newsrooms, and that represents a snapshot of the technological anxieties related to the replacement of human beings by machines. Now, unfortunately, we're only presenting in the virtual mode, and we don't have the time to do that. But if you're interested, then please go and take a look at the movie, which is really, really interesting.

While much has been made of its potential of freeing up journalists for more sophisticated workplace tasks, the introduction of new technologies in journalism has historically been met with a great deal of resistance, given that it challenges not only the boundaries of journalism, but it also invites journalists to reconsider their identities and ideology. Today, AI algorithms and machine learning are increasingly becoming part of newsrooms, influencing nearly every single aspect of journalism. We are therefore on the edge of an oncoming era of AI technologies, where tools not only impact newsrooms, but they can also have significant consequences for the public and the democratic role of communication itself. As news automation and artificial intelligence become pervasive in newsrooms, ethical principles such as accountability, transparency, reliability, and others lead us to carefully evaluate the ontology of these automated tools. But what do we mean exactly by ontology? We refer to how journalism is actually understood and what journalism stands for, especially in ethically related terms. We are referring also to the ontological boundary because when it comes to the evaluation of automated journalism, it lacks certain principles. Now over to you, Giulia.

Giulia Ferri Thank you. So as we have seen, journalists' relationship with technological innovations has been extensively investigated in recent years with different positions regarding the possibilities and the changes that arrive from their implementation. So we also know that the ethical and democratic principles are crucial in our definition of what is or should be journalism. So, for example, one of the first ideas and definitions that we have of journalism is to be a watchdog for the public service. So we have immediately an ethical and democratic consideration of it. But this is also true for automated journalism. In our study, we aimed precisely to answer these questions. So we focused on the relevance of these principles in the ongoing construction of the ontological identity and idea of how to make it journalism. So by analyzing the expert justifications and also considering other research on the subject, we tried to understand the value of these principles when it comes to justified implementation of artificial intelligence use in the newsroom, looking also to the challenges, and opportunities, and the problems derived by this implementation. The study was carried out as part of a project of journalism innovation in democratic societies, a comparative research of focus on the most important innovations in recent years. And the sample included five countries, Austria, Germany, UK, Switzerland and Spain. And for our analysis, we relied on the semi-structured interviews that we conducted in the first step of this comparative research, where we asked 20 experts which were the most important innovations of the last decade, from their point of view. Then we identified, and we filtered those innovations related to artificial intelligences, machine learning, or that included any form of algorithm. And eventually we conducted automatic analysis on the justifications of the expert.

Colin Porlezza Thank you very much, Giulia. So now when it comes to the findings, first of all, it becomes quite clear that the shared nomenclature is completely missing. Whenever experts refer to automated journalism, but also to artificial intelligence, there is neither a common vocabulary nor a common taxonomy regarding AI, which often results in quite a performative discourse about the potentials of AI that dilutes possible challenges that the technology entails. It is also interesting to look at the different legitimization strategies of the interviewees. When asked about the aim of using automated journalism, they may refer to the following three elements. First of all, automating so called "safe news." These are areas such as sports or financial reporting where no special expertise is actually required and where automation is quite easy to implement because it can also be based on a lot of structured data. This is actually in line with ongoing discourses in the news industry about the potential of natural language generation, especially in sports or business journalism. In addition, experts very often refer to the increasing efficiency that automated journalism entails by augmenting journalistic performance. It offers the possibility to not only produce more, but also to produce content faster and with far less expenses. And last but not least, automated journalism also offers the opportunity to personalize content in a far better way. It can be tailored to your target audience. It can also be used to moderate online communications by simply filtering out hate speech, for instance.

Now in the second batch of findings, we wanted to focus on what was actually absent of the experts' answers regarding automated journalism. So first of all, automated journalism can have implications that go far beyond the newsrooms, and these effects are simply not discussed. So economic logics that are driving the implementation of AI in newsrooms can actually have effects that transcend the boundaries of newsrooms. For instance, when it comes to news recommenders, we know that these can have a very detrimental effect on news diversity or even a healthy news media landscape, given that they are informed by economic interests and not primarily by societal perspectives. Another interesting finding is the fact that ethical issues, and this is the main reason why we referred to the boundaries

of the ontology when it comes to automated journalism, ethical issues seem to be quite irrelevant. Even if we asked our interviewees about inhibiting or challenging factors when it comes to the implementation of news automation, ethical considerations do not seem to play a primary role or they are not of primary concern. And also in this case, these findings confirm previous research. Last but not least, whenever it comes to artificial intelligence or automated journalism, as autonomous agents, that entail specific ontological issues in terms of an algorithmic reality construction, in terms of transparency, individualization, or a rising digital divide, are also not taken into consideration. So overall, these are the three main reasons why we think there are actually limits or boundaries to the ontology of automated journalism.

And this brings me to the conclusions of our presentation. So first of all, given the pervasiveness of AI in newsrooms, effects that, as I just said, go beyond the newsroom, should absolutely be tackled. Automation needs to be implemented carefully with regard to editorial tasks, and especially decision making should not simply be relinquished to machines only. Secondly, in our perspective, this kind of algorithmic construction of reality fosters debate-ification of journalism. So especially when it comes to the awareness of the consequences in terms of power relations within newsrooms, the autonomy, as well as the challenges to the authority of journalism, these are all absent, basically, in all of the legitimization strategies or the discourses brought up by the experts. The third point in the conclusion deals with the support of new editorial roles that we think are of primary importance. Journalists should not only play a role when it comes to shaping the design of automated tools, but they should also assume a new editorial role when it comes to retaining control, ownership, or even authority over data. So they should also play a far more active role within the newsroom when it comes to shaping not only the design, but also the pervasiveness and the impact that these technologies can actually have within newsrooms. And last but not least, we think in spite of the economic logics that are currently dominating the discussion around news automation in news organizations, we can see that also in many other reports and academic publications, we think that ethical issues should actually become a primary concern in news industry debates. Because particularly if we take into account issues of trust nowadays, ethical concerns are of primary importance when it comes also to this kind of self-reflection in terms of what are the impacts of automation and of artificial intelligence within newsrooms. Thank you very much.

Seth Lewis Thank you to our presenters. We have some time now for questions. I'll just say quickly at the outset that I would encourage you to look at the issue which is now online on the ISOJ website under research. I have a guest editor introduction there, where I talk about what I describe as the three D's of AI and news. I've never seen an alliteration I didn't like, so I couldn't help myself. And I talk about that these papers bring up really interesting questions about how AI has and will continue to influence the way that news is discovered, the way it's distributed, and the kinds of ways that it interfaces with democratic implications, as we just heard in this last presentation. So let's go back here to the talks we heard earlier, and maybe talk first about the AI for reporters tool. What are the likely next steps in that situation, and what might we expect to see in the future in that area?

Lindsay Grace Certainly, I think the core goal here is to iterate on the information that we've already received. Originally, the concept was to generate articles. What we discovered in talking with journalists is they actually wanted tip sheets instead, and now we actually know specifically what needs to be generated in those tip sheets. So the next two steps are to base it off of current data, so actually do all the mining for contemporary news, and also to generate tip sheets that are more accurate to journalists' needs.

Seth Lewis I wonder, though, looking at your slides, I mean, there were a lot of people involved, a lot of names, and everything, and clearly a lot of resources and so forth. I wonder if you might just talk a bit more about how feasible and practical are some of these types of developments? And then given that, what are we most likely to see in the next two, to three, or four years when it comes to these types of implementations?

Lindsay Grace Certainly, I think a lot of feasibility has to do with funding. So this project is sort of a derivation of the Digital Democracy Project from 2014 to 2018. And so a lot of the resources that were needed to run that would be similar to what you require for this particular project. So there's a lot of automation to scraping data and collecting this data, but at the end of the day, some of the fine tuning is a product of human interaction, as it often is. And so, you know, the future for this type of work is a blend of, I think, the work that computers do well and the fine tuning, the critical eye, that human contributions always offer. And I think that's a core concept in any of the work we do in AI, is to say that AI is not going to solve all the problems. It's simply going to augment our abilities and help us scale where things may be resource limited.

Seth Lewis So if we move from discovery to distribution and talk about your paper, Daniel. You know, I think what struck me was your slide at the end there. The kind of point being that there's a degree to which the autonomy of both journalists and of readers and users to shape the news agenda is compromised, or at least complicated by the role of algorithms. So talk just a bit more about that idea of like agency and autonomy and where those things figure into the shaping of the public agenda.

Daniel Trielli Yeah, no, absolutely. So when I started in journalism school, and when I started journalism, and I started working for a newspaper, the only person that I really needed to impress with my work so that it gets published and read, was my editor. Right? I just had to go through them to get my word out. And now what we see in the reality of journalism is that there is increasingly pressure from outside to shape your journalism, to frame it in certain ways. The most basic way would be SEO. These are choices that are made outside of the news organization. And these are things that journalists have had to navigate for more than a decade or so. I guess the point that we try to make is that there are limitations to what you might want to do. Autonomy also comes from within. So the same way the algorithm shapes, or might shape, or counteracts what the news value might be for the producer and even for the user, journalists themselves have to choose when to be shaped and when to counteract the algorithm as well. Right? And how to resist those things, or how or when to resist, and how and when to allow itself to be shaped by things. You know, we see different models of journalism that are not just focused on click-through rates anymore. We see, measurements of impact, particularly models that rely on funding sources that are not computational ads. So I think what we hope to nuance and complicate is how those processes can be affected and reshaped by journalists themselves.

Seth Lewis So we only have maybe two minutes, so we're going to wrap up with two very quick questions that are sort of related. One is where does the ethical conversation need to go in the news industry as it relates to AI and news automation moving forward? And I suppose the other kind of related question is what line of research in the kind of broad realm of AI, algorithms, automation, and news kind of yet needs to be more fully explored? So you can maybe take one or both of those questions. Maybe they're combined? I don't know.

Lindsay Grace I think as an educator, my emphasis is always on sort of the pedagogic, and our understanding of knowledge, and our understanding of what we consider to be, I think, an appropriate philosophical or ethical foundation for what we value and why we value it. So I think that's actually partly determined by the social pressures and the needs of our democracy. And so I think the smart thing here is often to have forward-thinking individuals, like the folks who are AI ethicists, helping us to guide that conversation because they have a broad perspective across a variety of genres.

Daniel Trielli I would add to that. I agree with that, but I would also add that we have been talking about values a lot in journalism internally, even before search and social media were a thing. So I think we can have those conversations as well within the realm of journalism. And I think in terms of research, we have a thing that is about to come out. I don't know when, because it's, you know, research and scholarly publications. So let's hope until the end of the year. We sketched out two ways of thinking forward in terms of research in regards to algorithmic audits of news information. One is about power, which is kind of what we did here. It's like who has the power to shape these things? And the other one is about values, right? What are the values of these actors? These are social, technical processes that involve multiple organizations with different perspectives. And I would be interested to explore what are the values of each of those organizations, particularly not only here in the United States, but also globally. When journalism in Brazil is also reliant or shaped by tech companies, then we see another division of geographical division and cultural division between countries.