Gemini Hegemony

ECIR 2026 — IR for Good Keynote

Dr. Madeleine I. G. Daepp Visiting Director of Civic Innovation, Public Democracy America
Senior Researcher, Microsoft Research

SLIDE 1: Today I want to tell you about a plant physiologist named Arthur Galston. Like a lot of researchers in the mid 20th century, he wanted to feed the world.

SLIDE 2: ARTHUR GALSTON WANTED TO FEED THE WORLD. AND HE DID. In his doctoral thesis he studied a chemical called triiodobenzoic acid, or TIBA for short. He found that it could speed up soybean flowering but, in high concentrations, it could also defoliate those same soybeans. For Galston, it was a promising fertilizer. Now most graduate students are excited if anybody reads their thesis. Years later Galston would be delighted to learn that a company patented and commercialized it with farmers, increasing soybean yields by almost a third in some places. But researchers at the U.S. Military also read his thesis. And they used it to create herbicides that they sprayed across Vietnam to defoliate trees so they could better see their targets.

SLIDE 3: HIS WORK ALSO LED TO AGENT ORANGE. Galston's work became a core component of Agent Orange.

SLIDE 4: "I USED TO THINK THAT ONE COULD AVOID INVOLVEMENT…" Galston spent the late sixties speaking up about the chemicals' harms. In 1970 he testified to congress that Agent Orange could harm humans. His testimony led the Department of Defense to run laboratory studies showing it could cause birth defects in rats. The next year, Nixon ordered its use to be stopped. But a huge amount of damage was already done. For the rest of his life, Galston would teach bioethics classes and warn other scientists.

> "I used to think that one could avoid involvement in the antisocial consequences of science simply by not working on any project that might be turned to evil or destructive ends. I have learned that things are not all that simple, and that almost any scientific finding can

> be perverted or twisted under appropriate societal pressures. In my view, the only recourse for a scientist concerned about the social consequences of his work is to remain involved with it to the end."
>
> — Galston, Arthur W. *Annals of the New York Academy of Science* (1972)

Simply wanting to do good, simply building towards good ends, is not, in fact, how good gets done. Instead, these grand visions, from feeding the world to organizing the world's information — these visions are, more often, how the very smart and technically adept people who want to do good in the world get taken advantage of by the less skilled but power- and money-hungry.

SLIDE 5: IF WE WANT TO TALK ABOUT IR FOR GOOD, WE NEED TO TALK ABOUT GENERATIVE AI. I know nobody wants another AI talk. But if you want to take seriously the idea of doing IR for good, you will need to trace it to its end uses. Because so much of how generative AI is being used right now depends on information retrieval. And so, I promise you that I will talk about IR, but first we do have to talk about AI.

SLIDE 6: DRAW A UNICORN IN TIKZ. Three years ago now, I was in a small seminar room with a few dozen Microsoft researchers to see a demonstration of a novel technology that, we were told, would change the field of computer science. We would have to redesign the entire stack from top to bottom.

That technology, of course, was GPT-4. And of course you all know, now, what I learned that day but just so we are on the same page: it was a paper titled "the sparks of artificial general intelligence", now colloquially called sparks. That paper has been much maligned, and we can talk about its issues, but in my opinion the really important finding was this unicorn. You see the thing about this unicorn, is that it was made in tikz, the graphics plotting language in Latex.

Now the thing you need to know about Tikz is that it's absolute trash. As academics of course one way we communicate our seriousness is by making our work ugly, and so LaTex is distinctive for its deep academic ugliness. And so anything that looks even a little bit like a real unicorn is frankly amazing.

But more important than that is how this GPT-4 unicorn compared to what you could get from GPT-3.5-Turbo or what we all knew as ChatGPT.

SLIDE 7: IS MORE DIFFERENT? The idea at the time was that what we were seeing was emergent laws of scale, that remarkably simple systems can suddenly do quick complex things when they get big enough. And for the haters in the room it's maybe helpful to take a sociotechnical view of the thing, and to remember that in big tech everyone has had their careers and fortunes entirely shaped by Moore's law, the observation that the number of transistors in an integrated circuit doubles every couple of years. And it's not clear if Moore's law is actually a law or a self-fulfilling prophecy, a sort of social phenomenon where, in convincing everyone it was a law, it made itself true.

So any time you can show people in the bay area a log-linear plot, they get very excited and they throw money at you, and in turn that can make the thing happen.

A quick caveat: it turns out a lot of these step-function curves are artifactual, but what *has* held up is that there are these real regime changes where as the models get bigger and new methods come online, they really can do new things. I know a lot of wonderful folks who resisted redirecting their research towards it, who had their own problems from esoteric mathematics to developing new carbon-capturing concrete, where they were making progress and wanted to keep working. But I decided to switch up my whole research agenda.

Sure, other problems are important. And probably there was some of that arrogant tech grandiosity you get when you grow up reading Feynman and Asimov, and you've desperately hoped we might one day find ourselves in the right room at the right time to play some tiny role in some big moment. But to ignore AI, I thought, would be rather like being a graduate student at UCLA or Utah and ignoring the ARPANET.

SLIDE 8: 2024: THE BIGGEST ELECTION YEAR IN HISTORY. I should tell you that at the same time, I was reading quite a bit about how 2024 was going to be the biggest election year in history, with more people eligible to vote than ever before. Now, that happens every time India votes — but the year was also special because it was the alignment of a set of 5, 4, and 3-year election cycles that happens, of course, only every 60 years, and so there were over 74

countries voting with a total affected population of about 4 billion people. All told, according to the International Institute for Democracy and Electoral Assistance (International IDEA), about 1.6 billion ballots would be cast.

So, in the interest of seeing the tech through to its end uses, I pitched a research project that would take me to the first democracy to vote in 2024, Taiwan and the largest democracy, India.

Taiwan has more external information operations than any other place in the world, according to the Swedish Varieties of Democracy Institute, and of course the major actor involved has its own models, so I talked to the factcheckers, journalists, cybersecurity experts, and officials who were effectively democracy's canaries.

SLIDE 9: THE DEEPFAKE FEAR. A big concern, for the Taiwanese defenders, was deepfakes — deceptive and adversarial depictions of people and events meant to convince their viewers of something that did not happen. As Stella Huang, a journalist at a major Taiwanese news outlet described it:

> "So we're very afraid, and international media are very afraid, that on the eve of the election some AI-generated media will come out and we won't have time to clarify, to debunk."

This was also a major concern of the defenders in India. And it was frankly the top concern of international media. But it's not what happened, not in either place.

SLIDE 10: WHAT WE ACTUALLY SAW. This is a clip from a Bollywood movie. But they've face-swapped in the heads of India's two opposition parties, Rahul Gandhi and Akhilesh Yadav. This isn't a deep fake, right? This is just the political cartoon of the AI era. This is satire.

As the technology reporter Nilesh Christopher described it:

> "Basically everybody thought deepfakes were going to have this scorched earth future, but it's effectively being co-opted only for two purposes. One is voter outreach and another one is political memes, for shits and giggles."

I do think there's this huge problem in the journalistic space where everyone calls everything a deepfake. So what we saw were all of these promotional, really laudatory uses. These have been called soft fakes, because they were most effectively used in the Philippines, where Prabowo Subianto, a former military general implicated in the murders of hundreds of dissidents, successfully rebranded himself as "cuddly", softening his image.

There were also cases of voice dubbing across languages with the candidate's blessing — authorized fakes — and on the other end, a genuine deepfake of a US senator supporting a Taiwanese party. But it didn't take. Why?

SLIDE 11: SOCIOTECHNICAL SYSTEMS HAVE SOCIAL AS WELL AS TECHNICAL DEFENSES. Now the thing about deepfakes is that they presume that a technical advance, necessarily, constitutes a real-world advance. But you don't have to have a PhD in sociotechnical systems to know that's not true. At my various day jobs I've seen dozens of brilliant technical advances that languished because the market wasn't ready, because people just didn't use or want it. And when we talk about adversarial tech like deepfakes, it's important to remember that sociotechnical systems have social as well as technical defenses.

Let me say that again. Social systems have social defenses.

SLIDE 12: SOCIAL DEFENSES. India has mostly internal influence operations. And so the nice thing about that is you can actually interview some of the folks engaged. That is, in India, in addition to talking to factcheckers and journalists and activists, I also talked to political consultants from multiple major parties, to social media marketers and advertisers and one very interesting man who called himself "The Indian Deepfaker". His name was Divyendra Singh Jadoun, and he's the founder of AI startup Polymath Synthetic Media Solutions. He and other technologists published an ethical compact, pledging to maintain transparency in AI's deployment, and he decided to include a visible watermark on the content he created. In India, where representational misuse could lead to defamation charges, making AI's use obvious was a strategy for reducing the legal risk to creators.

Now this is trickier in Taiwan where a lot of actors are external, but they do collaborate with internal influencers and in Taiwan, as in India, there was a very high-profile legal case leading up to the election with big fines and criminal charges for a deepfake creator.

The influencers we talked to also worried about the reputational hit that would take if they circulated a deepfake. For example, I spoke with a social media marketer, Balram Vishwakarma, who said that he would take a big reputational hit. He had gotten his start by building a successful account, Andheri West Shitposting, that gave him cred when he reached out to others. If he posted something shady, he could lose years of building his audience overnight. For influencers with consistent identities and reputations they've built over time, there are real costs for misuse.

And finally but maybe most importantly, there's real importance to making sure the deepfake is about something actually believable. When somebody puts a fake video or image online, there's a whole social sensemaking process that happens. Here's Eve Chiu, the CEO of the Taiwan factcheck center, explaining how she knew a video of a US senator supporting a Taiwanese party was not real:

> "By our journalistic expertise and geopolitical knowledge. We know if a U.S. Congressperson says things like that, there will be an international outcry…"

And of course she's a trained factchecker. But everyday people do this too. For example, you might remember that time Kate Middleton attempted to pass off a fake photo as a real appearance, or when Drake dropped a track with a Snoop verse. Royals watchers and hip hop fans aren't trained in OSINT or factchecking — but in those social ecosystems, they didn't stand a damn chance.

More often it's not the social factchecking but just the fact that nobody cares, that nobody is checking TikTok to see U.S. senators speak. Most people are on the internet to see cat videos. Nobody cares — unless it's the night before an election or the morning before a big storm or right after an invasion.

SLIDE 13: HOW EASY IS IT TO GENERATE NARRATIVES? So if the impact of AI isn't in deception, then what is it? I want to make the case to you that it's really about efficiency gains for bad actors. How easy is it, really, to generate this kind of content at scale and across modalities?

And what's also easier than ever is to do this across new languages and cultures. This was my biggest concern from Taiwan and from India, that when narratives did catch on they caught on in impeccable traditional Chinese, in Marathi, in Telugu. The content farms and dark PR firms that used to work in big languages and in countries where the oligarchs could afford their services seem to be really branching out.

And so at all of the frontier model companies there are these groups called red teams that get the models two weeks before the rest of the world and try to be the last internal line of defense. And so I wanted to come up with a way for the red team to look at this problem of scaling information manipulation across languages and cultures. The problem is that a lot of the red teams tend to be western educated industrialized rich developed, so they don't really know what narratives are going to hit in other places. And you can hire people from anywhere in the world, gig workers, but I'd argue that writing a really spicy disinfo narrative is actually hard.

So our insight was that you could just take real examples of narratives that were hitting in order to know what it was you needed to defend against. And at the time there were still people collecting those narratives. Over 70 countries had their own factchecking outfits. Now the situation has changed, but you don't need to use factchecking claims, you could use community notes or OSINT reports or what-have-you. The important insight is just to use real narratives from real places in as many languages as you can. We clustered those claims into dominant narrative themes, extracted knowledge graphs encoding the key entities and relationships in each narrative, and used those graphs to augment an attacker LLM.

We collected claims from fact-checking organizations in English, Hindi, and Spanish across India and the United States. In India, the prominent narratives centered on religious conflict and PM Modi; in the US, on COVID vaccines and the 2020 election.

SLIDE 14: GPT ATTACK SUCCESS RATES EXCEED 80%. What we found, off the shelf models, was attack success rates exceeding 80% on the GPT models across every language-location pair we tested.

Now I will say GPT-5 does better on this, but there's a principle I want you to hold onto: you are only as safe as your least safe model. Not only can adversaries switch to whichever model has the fewest guardrails for their particular language and topic, but once you have a model that generates attacks, my colleague WeiWei Yang has shown you can use it to circumvent other models. One bad model ruins the bunch. It's a sort of reverse herd immunity effect.

SLIDE 15: THE CHALLENGE IS IN EFFICIENCY GAINS.

> "It's a DDoS. It's a denial of service on the democratic language, which is always so limited before the election anyway."
>
> — Audrey Tang, then-Digital Minister of Taiwan

But what is the avenue for impact of all of those narratives? Well, there's a phenomenon called the search engine manipulation effect.

SLIDE 16: THE SEARCH ENGINE MANIPULATION EFFECT. In a study published in PNAS in 2015, researchers showed that reranking search results so that those favoring a particular candidate were at the top had a large effect on the perceptions of prospective voters.

Now the researchers ran multiple versions of the study, most of which were the sort of thought experiment-y type "look at these candidates in an election of a country you aren't a citizen for" setups. But they also ran it in the real world. In 2014, in India, they recruited 2,150 actual voters across 27 states in the 2014 Lok Sabha elections. And first they had these voters read biographies and rank the candidates. And then the voters got fifteen minutes to use a search engine, kadoodle. And this was a double-blind study, so the participants didn't know it but they were randomly assigned to a kadoodle in which search results were reranked favorably for one of the three candidates (versus no reranking in the control group).

Now for those of you who care about research ethics, they argue this is ethical because its such a small number of people relative to the scale of the electorate. It's wild their IRB approved it, but it did and so we have to take it seriously.

What you're seeing is a marquee finding from that paper, which I recreated with AI for all of you.

And what you see here is that, in every case, the biased search rankings led to a significant increase in the fraction of votes that the relevant candidate received. Now these effects are small, on the order of 0.2-0.4 points on a ten point scale. But all effects are big effects when you've got a population of 1.4 billion. The authors estimate that this effect can shift the preferences of undecided voters by 20% or more — magnitudes that, at population scale, could meaningfully change election outcomes.

In a follow-on study published in CSCW in 2017, they do 3 more experiments with 3,600 people across 39 countries, and find the effect and its magnitude are attenuated but still present even when you alert people to the bias.

To be clear I do not think that Google is doing anything shady with their search rankings. Indeed, independent evaluations show that they do a good job of defaulting to authoritative content on major elections. The problem isn't that they do but that they could.

And the other problem is that all those narratives you just saw? This is one of the things they are trying to move. If it's the night before an election or the day of a big hurricane or a bombing in Tehran, people are going to use search to figure out what to think. Those are the moments that narratives get set.

We are already getting so many studies about whether AI deceives, whether it persuades. And they are all testing those effects directly when in real life all of these effects get mediated through search. And this happens, these days, on a number of platforms. Younger users are searching on TikTok, mostly, but 90% of text-based search still goes through Google.

SLIDE 17: FROM SEARCH TO SUMMARIES. A few years ago I was studying what I called the emerging AI divide, or the fact that there were all these people on the west coast of the United States talking incessantly about this thing called ChatGPT but in rural regions in the American south and our industrial midwest the only time anyone talked about AI was if they were getting nudified or pig butchered by it. And at that time only about 20% of people had tried an AI chatbot. A year later a majority of people had tried it. And then overnight, tracking whether people were using generative AI became entirely obsolete, because Google rolled out Gemini summaries and suddenly basically everyone, everywhere, with internet access was using generative AI overnight.

There's a great new study by Sinan Aral and coauthors from MIT, who ran 12,000 queries in 2024 and again in 2025 to track how AI summaries were being rolled out. They found that AI overviews showed up in just 7 countries in 2024 versus 229 in 2025. So what does that mean for our elections? Because Google has 90% of search, meaning that Gemini is now the voice behind AI summaries all over the world.

So what does it mean, for the search engine manipulation effect, when we have summaries instead of pure ranked lists?

Well, I can't tell you. You all have good methods for measuring bias in ranked lists — Abolghasemi's term-based approach from last year's ECIR is an example. But we don't yet have equivalent methods for summaries, and we especially don't have them for low-resource languages. Nobody has yet, to my knowledge, tested the SEME effect in summaries.

SLIDE 18: AI SUMMARIES CAN SHIFT OPINIONS. There is one interesting study by Yiwei Xu, Saloni Dash and others out of UW showing that, in a pre-registered RCT, AI-generated summaries shifted user positions on contested issues in the direction of the summaries' bias, with stronger effects the more prominent the summary's page placement.

In this case the control *was* just seeing the ranked lists and the bias was in the summarization, but the search results are of course going to be a mechanism through which this occurs. There's some early work, for example, that shows how search order shapes results. In particular, Jing Ma at the University of Zurich showed that even when prompted to produce neutral summaries, given three ordered search results Gemini 2.5 tended to be most semantically aligned with the first-seen article it was given to summarize. So traditional SEO still holds power in the GEO age.

SLIDE 19: FROM SEARCH AND SUMMARY TO EVERY INTERACTION. But now let's escalate a notch more. Let's put the same AI model not just in summarizing search, but in every document you write and every email you send. What happens then? Well, according to a wildly important paper that just came out in Science Advances, Williams-Ceci and coauthors find that LLM autocomplete tools can have statistically significant effects on people's actual political positions. In a beautiful set of preregistered randomized experiments, they had people answer a political survey and then, several weeks later and without knowing the studies were connected, provide written answers to questions in a setting where they could randomly be assigned a bias or neutral AI writing support tool. It turns out that the result was a shift in opinion on the issue, even when the subjects got a warning or a debrief.

This is a very important result, and also a reminder that if you do not want Grok to have direct access to your precious brain you need to get off X. Because again, even when people *knew* that the AI was biased, the effect still held.

SLIDE 20: THE LESS YOU'RE IN THE TRAINING DATA, THE WORSE AI SERVES YOU. And again this has yet to be properly studied, but it's probably going to be so much worse if you're not WEIRD or more specifically WAT or West-coast American Techie, because when you go to

look yourself up you get this weird version of you that somebody else has written. My friend Rida Qadri has written about the ways AI reproduces an outsider's gaze; my participants in India described it as a sort of Hollywood view of the Bollywood view of Indian culture.

In fact in my own interviews there was this fascinating rich-get-richer effect, where the more well-represented someone was in the training data the more useful they found the tools and, in turn, the more they were represented in the data.

SLIDE 21: GEMINI HEGEMONY. So this brings us to the idea of hegemony. And I know you are all more comfortable with equations and formulas but bear with me as I get a little bit theoretical here. So, historically, hegemony is the dominance of a particular city state. But the Italian theorist Antonio Gramsci talks about cultural hegemony, about the way that dominance is maintained through cultural and ideological consent. The hegemon, in this context, shapes the internal politics and character of the subordinates.

For example: I flew here from Seattle and did not learn a single word of Dutch. *Dank je wel*, maybe. You all speak English fluently. Why? Because American cultural hegemony shaped what "participating in international academia" requires. And that means that in order to participate in international academia, you have to expose yourself to American cultural norms. Which brings us to an insight from the communications scholar Andrea Mayr, that "as a practice of power, hegemony operates through language." What she's saying is that language shapes how we describe and interpret social realities. And our AIs are about to shape and reshape most of our language.

Again, I'm not saying that Gemini or Grok or my own company's Copilot *are* changing political views at scale in any intentional way. Well, somebody should probably check on Grok. What I'm saying is that they are reinforcing hegemonic discourse and if you are a user who is not well served by that discourse then you are about to find yourself in a pickle. What I'm saying is that hegemony operates through language, and Gemini is about to mediate the vast majority of the online language people in over 229 countries encounter.

That's what I mean by Gemini hegemony. It's no shade to Google. It's a call out that a single model, trained on a particular corpus, representing and reinforced by a particular worldview, is becoming the default lens through which billions of people understand the world and, in seeking to understand the world, come to understand themselves.

SLIDE 22: GEN AI WILL DEPEND ON RETRIEVAL. But you all are not just users. You are also information retrieval researchers. And the thing about these Gemini summaries is that of course they actually lie on top of search. Several big breakthroughs in gen AI over the last couple of years tell us there is about to be a lot more retrieval in everything. Recursive language models, memory systems, and grounding all rest on retrieval. And retrieval, at its core, is a pluralistic technology, a technology that only works well when there's diversity.

SLIDE 23: IR IMPLICATION 1: BUILD SEARCH ALGORITHMS THAT FOSTER EPISTEMIC PLURALITY. So what does that mean in terms of a research agenda for IR for good?

SLIDE 24: IR IMPLICATION 2: DESIGN FOR CONTESTED ECOSYSTEMS. Now let me also be clear that the issue is not *just* with the providers. It's also with the data that goes into them. These are the second-order effects of contemporary sociotechnical systems. Or if you want to frame it from in security lingo, this creates an attack surface.

If you want to get into the Gemini summaries, if you want to get into the training data, what you should be doing is perturbing all this stuff at scale. And I hate to tell you this, but the global networked kleptocracy is throwing all kinds of money to do exactly that. And meanwhile, the NYT and Elsevier, they're actively taking themselves **out** of the training data. If you think social media has brought on an anti-fact anti-science age, honey, you ain't seen nothing yet. So in keeping with Galston's insight that you need to trace not just what your tech could do but what it actually does, the IR for good research agenda will also need to recognize that you are building for a contested ecosystem where you need to think about second order and social effects.

SLIDE 25: THE ARPANET BUILDERS. Let me tell you about some people who were a little different from Galston. For one thing, they were a bunch of people, not just one man. And they were hippies, eager and idealistic grad students at Utah and UCLA who didn't yet know what this new technology they had gotten their hands on would be able to do. When the military generals paid for these young professors and graduate students to build the ARPANET, they wanted something hierarchical, a military technology. So how did we get what we got? Well, the computer scientist Taylor said, of his approach to — again — this technology literally funded by the defense department:

> "My bias was always to build decentralization into the net. That way it would be hard for one group to gain control. I didn't trust large central organizations."
>
> — Bob Taylor, Director of the Information Processing Techniques Office, ARPA

Later on ARPA's then deputy-director told Stephen Crocker: "I was on top and you were on bottom, so you really had no idea of what was going on and why we were doing it." But what Crocker, who had been a grad student on the project, said was: "I was on the bottom and you were on top, so you had no idea of what was going on or what we were doing."

The future belongs to the builders, which is to say it belongs to all of you. What would it mean, in setting up the next version of grep or the next solution to grounding or whatever mess memory ends up being, to say intentionally: what is the pluralistic solution here?

The dream of those early internet founders was to connect everyone to the glorious richness of the whole world. What if we take those core tenets that are already so ingrained in internet and technology culture, the ideas that so effectively drive Wikipedia and Linux and Git, and turn them into the foundational frameworks of AI?

So I have to tell you, I finished all this research in a pretty dark place. But a few months later, I went to a talk Audrey Tang gave in Seattle. And they said something really interesting. Of the Taiwanese civic tech and gov movements they said you know protests are important. But protests only show people what you are against. Demonstrations build the future you are **for**.

SLIDE 26: IR IMPLICATION 3: DEMONSTRATE ALTERNATIVES. So what are the demonstrations that we need from the Information Retrieval world?

What's the truly pluralistic search algorithm? What's the ranking and recognition method that reinforces diversity and doesn't collapse into hegemony? I have no idea. But somebody in this room might. And I'd rather you build it than wait to see what the next Gemini update does. I don't know yet, what kind of future we will find in the AI age. But instead of waiting to see what someone else will build, what we need are some damn demonstrations.