RESEARCH STATEMENT

AMY X. ZHANG

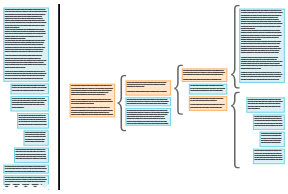
Discussions online are integral to everyday life, affecting how we learn, work, socialize, and participate in public society. Yet the systems that we use to conduct online discourse, whether they be email, chat, or forums, have changed little since their inception many decades ago. As more people participate and more venues for discourse migrate online, new problems have arisen and old problems have intensified. People are still drowning in information, with few mechanisms for managing or synthesizing large volumes of discourse. Along with scale, users must now juggle dozens of disparate discussion silos, spread out across different apps and websites. Finally, an unfortunately significant proportion of this online interaction is unwanted, untrustworthy, or unpleasant, with clashing norms leading to back-and-forth bickering, people getting harassed into silence, and misinformation running rampant. Left unchecked, these problems have far-reaching harmful effects on our society.

My research in human-computer interaction is on reimagining outdated designs towards **building novel online discussion systems** that fix what's broken about online discussion. To solve these problems, I develop computational techniques and tools that **empower users and communities to have direct control over their experiences and information**. These include: 1) summarization tools to make sense of large discussions, 2) annotation tools to situate conversations in the context of what is being discussed, as well as 3) moderation tools to give users more fine-grained control over content delivery.

In my work, I conduct in-depth qualitative inquiry and large-scale quantitative data analysis towards understanding issues that users have with online discussion, before developing new computational techniques that meet those user needs. Finally, I design, build, and deploy systems that use these techniques to the public in order to achieve real-world impact and to study their use by different communities. Given the public-interest nature of my work, I also pursue broader societal impact through outreach to industry and the public and collaboration and coalition-building with diverse parties. I have successfully collaborated with over 10 outside groups, including within industry, other universities, nonprofits, journalism groups, and civic organizations.

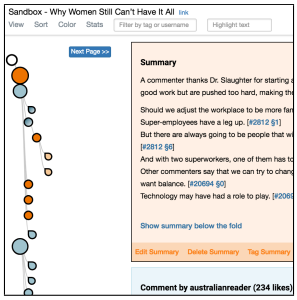
MAKING SENSE OF LARGE DISCUSSIONS

While much of the world's factual information is readily accessible via a search query today, there is still a wealth of experiential, contextual, and opinionated information in first-person accounts and discussion online. Unfortunately, this information is often embedded within sprawling conversations that are near-impossible to follow, sift through, or comprehend. The first contribution of my research is a set of systems that give users the tools to organize and summarize large discussions. Because discussions are often too large for any one person to distill, the techniques I develop, while applicable for individual use, allow summarization to scale with the size of discussion by enabling collaboration.



Left: Discussions can be long and hard to read. **Right**: Recursive summaries enable progressive hierarchical exploration.

Wikum [1] is a tool that scaffolds the complex task of summarizing and organizing a large threaded discussion. Wikum instantiates a crowdsourcing technique I designed called *recursive summarization*, where users build summaries of small sections of the discussion, small sets of those summaries are then aggregated and summarized, and so on until the entire discussion is summarized. Wikum also incorporates techniques from visualization and machine learning to aid users, such as a directly-manipulable tree visualization of the discussion, clustering and tagging suggestions to find related comments to group, as well as automatic summarization algorithms to assist with summary writing. The result of the workflow is an explorable *summary tree artifact* that bridges from a high-level wiki summary to more focused summaries to the original back-and-forth forum discussion.

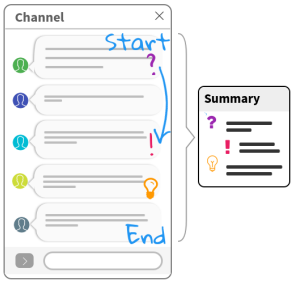


Visualization of a partially summarized discussion in Wikum.

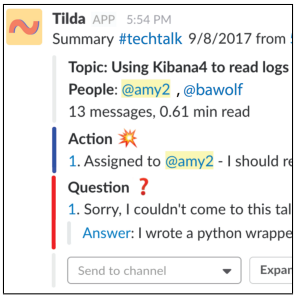
Through a collaboration with the Wikimedia Foundation, Wikum is in use by top Wikipedia editors who often struggle to comprehend and resolve contentious content disputes. In a comprehensive analysis of formal deliberations on Wikipedia,

we found that a third go unresolved due to factors like long and complicated back-and-forths that are hard to follow [2]. Using Wikum, editors were able to reduce their cognitive load when summarizing deliberations, transforming a previously solitary task completed in several intensive hours into one that can easily be distributed across multiple sittings and editors. Wikum was also selected by the Medialab Prado of Madrid as a showcase tool for citizen democracy, and early studies have been conducted towards the use of Wikum by citizens in city platforms such as Decide Madrid to summarize discussions about city-wide proposals.

In the case of more real-time chat, popularized in the workplace with tools like Slack, catching up on missed conversations can be a struggle. Through interviews with people who use group chat for work, I learned that scrolling is the dominant strategy for catching up, and that making sense of what was said is difficult due to the lack of information signals or structure to differentiate chat messages. I then built *Tilda* [3] in collaboration with Microsoft Research, a tool that provides affordances for rich *markup* over chat with information pertaining to the structure, role, and importance of messages. Examples include adding major discourse acts, such as "question" and "answer", linking from one message to another, and delineating separate conversations. Because much conversational context is lost after a conversation is over, Tilda builds in lightweight techniques for *in situ markup* integrated within the chat application, including both text commands in the chat dialog box and direct manipulation via emoji reactions. The markup is then used to automatically construct short summaries of conversations that let new readers quickly get an overview and dive in to the original chat messages that are of interest. In field studies with real teams on Slack, including two software startups, a research lab, and a news publishing team, Tilda was found to be effective for catching up on chat. This work received a Best Paper award at ACM CSCW 2018.



Marking up chat to generate lightweight summaries.



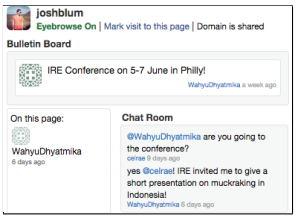
A discussion summarized in Tilda with discourse acts marked.

PUTTING DISCUSSIONS IN CONTEXT

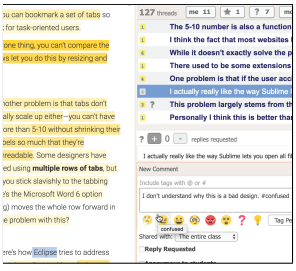
Too often, online discussions live within dedicated silos, completely separate from the thing that they are discussing. Student discussions on Piazza cannot reference relevant course materials [4]; developer discussions about code on Slack are completely separate from the actual code base [5]. This can lead to a disjointed discussion space overloaded with different topics, making relevant discussions difficult to discover. I explore a series of designs that make it easier to have discussions *anchored* to a piece of primary content so that context is preserved and discovery is easier.

One such system is *Eyebrowse* [6], a discussion layer over the entire web. Users designate the online spaces they consider public—spaces where they can "bump into" friends, see where crowds and trails form, leave messages, and have impromptu conversations, much as they might in physical public squares. Conversations on Eyebrowse become attached to the page where they occurred, allowing future visitors interested in the same content to discover relevant conversation. One important consideration in social applications for browser activity sharing is privacy. In Eyebrowse, I designed a *domain-level consent model* where users build up a whitelist as they are browsing of sites that they consider public. Through public consented sharing, Eyebrowse makes possible an ecosystem for developers to build tools that give benefits back to the public instead of private companies as is the case today.

How might annotated discussions scale when there are hundreds on a single page? I explore this question by building off of the *Nota Bene* (NB) system [4], a widely used tool for having anchored discussions "in the margins" of online textbooks, where I



A serendipitous conversation between friends on a webpage in Eyebrowse.



A comment expressing confusion in the margins of a textbook in NB.

developed a feature for users to self-tag their comments with hashtags signifying their emotional state, such as "confused" or "curious" [7]. This information allows instructors to find the threads that need a response as well as understand at a glance what sections of the textbook caused confusion. Since the feature has been released, over 150,000 comments have been left behind containing hashtags by nearly 6,000 students, including in classrooms with enrollments of over a thousand students. In addition to providing a useful signal for instructors, this rich dataset allowed me to train models to classify untagged comments with their emotion or predict which sections of a textbook are likely to be confusing.

CONTROL OVER MESSAGE DELIVERY

Whether receiving or sending emails or posting or viewing posts on social media, users often have little control over the mechanisms of that delivery. As a receiver, it can be hard or impossible to control what kinds of content to receive or when or how to receive them. On the sender side, messages are often all or nothing—either everyone gets them immediately or not at all. A prime example is the humble mailing list, a widely used system that has barely changed in 40 years. From studies of both workplace and social mailing list communities, I found that, paradoxically, people often wanted more substantive discussion but were themselves too shy to post for fear of spamming. I also uncovered tension between members due to conflicting ideas about appropriate behavior, partially influenced by how they configured their mail delivery. Motivated by this work, I developed *Murmur* [8], a reimagination of the mailing list system that allows members to more finely configure what messages get delivered—for instance, by following threads, individuals, or topics of interest. Conversely, receivers can block topics or only get the initial message of threads, while senders can target to a specific audience or slow a message's propagation. By providing a way for both senders and receivers to fine-tune their delivery, messages can collectively go to only those who want to receive them. Following this work, we have conducted studies of email more generally to understand the many customizations users want regarding delivery, notification, and presentation towards developing a simplified domain-specific language for expressing email automation [9].

An extreme example of losing control over delivery occurs in the case of online harassment, where people with an intent to harm flood a recipient's inbox with hurtful or disruptive messages. Here, too, users need more tools to maintain control over what, how, and when to receive content but may still feel overwhelmed and vulnerable working alone against determined harassers. I led an MIT M.Eng. student in the development of *Squadbox* [10], a tool that introduces the privacy-sensitive technique of *friendsourced moderation*, where a recipient of harassment can automatically forward suspicious messages to friends who then moderate them according to the recipient's wishes. Unlike spam, harassment is defined in many ways, and recipients have differing preferences on how to deal with harassment. As a result, Squadbox is designed to be fully customizable by the recipient. Since its release this year, Squadbox has been profiled on BBC and Channel 7 News Boston TV, as well as in articles by ABC News, the Verge, Business Insider, New Scientist, and Fast Company. In outreach on this topic, I have also spoken about Squadbox and content moderation more broadly at non-academic events like Mozilla Festival and RightsCon, and am currently an advisor to OnlineSOS, a nonprofit countering online harassment.

Ways to use Squadbox: **1)** auto-forward certain messages from one's inbox to friends, or **2)** create a public-facing moderated account.

An equally alarming example of loss of control occurs with the problem of misinformation, where users have little ability to distinguish between trustworthy and untrustworthy content in their social feeds or filter false content out. While AI has been proposed as a solution, opaque and uninterpretable machine learning models for filtering misinformation are concerning for free speech and user autonomy. In the past year, I became a founding member of the Credibility Coalition and a member of the W3C Credible Web Community Group, two international groups developing transparent and interoperable standards for news credibility. As part of this work, I led a research study with 13 co-authors on defining and annotating credibility indicators in news articles [11]. My study led to a partnership with Google Jigsaw and funding from Google, Facebook, and the Knight Foundation. I have spoken about this work at Mozilla Festival and the International Journalism Festival and was recently covered by Poynter Institute as one of six academics on the "frontlines of fake news research."

RESEARCH AGENDA

My goal is to build technology that empowers users in their everyday interactions with other people and with discussions online. In this section, I outline my future research directions that I am the most excited about.

Novel Interfaces that Harness Information within Discussions. The techniques introduced in Wikum and Tilda are just two of many ways that people can provide signals to synthesize and add structure to discussion. As an example of another signal, thousands of readers may explore the same forum, creating paths of interest through different threads, yet each newcomer must start again from scratch. Harnessing both active and passive signals will be necessary to build web-scale systems for synthesis and exploration of discussion. For instance, what is the best way to build the Wikipedia of opinions or the Google of public forums?

An additional follow-up question is how such systems would look and behave. While discussion threads have little variation in appearance today, the actions that people take while conversing can vary greatly. Building on work I conducted on automatically characterizing common discourse act chains [12], I aim to explore new representations of discourse beyond simply threaded and non-threaded, towards a typology of discourse structures. I am also interested in exploring beyond textual conversations to develop techniques for synthesis of spoken communication, given improvements in speech-to-text technology. This includes phone, video, or in-person meetings as well as larger collocated gatherings, where many discussions may be going on simultaneously.

Towards Productive and Civil Discourse. While Wikum supports understanding of discourse after the fact, and Murmur and Squadbox support moderating out unwanted discourse, I seek to also design systems to improve the quality of discourse for discussants. One important aspect is productivity or how to move beyond redundant or excessive bickering, as I found this was a hindrance for peer production communities like Wikipedia [2]. Moving forward, I plan to explore how discussion systems can give participants a sense of progress, even when there is no tangible output. One way could be to treat the discussion artifact itself as an output or an input into another artifact, as opposed to something discarded once over.

Another important and related area of improvement is how we can increase empathy across sides, particularly given the prevalence of polarized discourse readily observed online today. I hypothesize that an important aspect of breaking through echo chambers is *framing*, particularly the underlying moral frames behind arguments. Early experiments I have conducted in this area provided evidence that getting discussants to share their moral values with others increased empathy towards the other side. Work I have done to externalize, detect [13, 14], and visualize [15] moral values could give people a better frame for interpreting different opinions. Looking forward, I will explore how these concepts apply to interfaces for discussion, news sharing, or recommendation.

Human-AI Governance of Online Content. Squadbox presents an example of *networked moderation* for when centralized platform moderation is deficient or not a good fit. Today, we are rapidly moving towards a world where online discourse is governed by a few major platforms, a concern when it comes to the health of public discourse on the web. In addition, as platforms struggle to keep up with moderation demands and increasingly rely on automation, it will be important to consider how users can still have agency. A focus of mine starting this year with my fellowship at the Berkman Klein Center at Harvard University is on alternative modes of online governance that allow individuals, networks of trust, and groups and organizations to have a say. Beyond machine-interpretable signals, I believe we need human-interpretable tools for filtering untrustworthy content, and that go even further to let people modify, collaborate on, or create filters. This means that future governance models will need to examine the best ways for humans and AI to collaborate. For instance, work I conducted at Google explored human evaluation of clustered search results [16]. As more machine learning is deployed to real-world applications, I plan to explore new human-centered ways to evaluate the performance of models.

Conclusion. The decisions we make now in the design of online discussion systems will have significant ramifications touching almost every aspect of society. In my research, I design and build systems that give users the power to have and make use of online discussions in new and better ways. In the future, online discussions will not simply be a poor simulacrum of in-person discussions but will support a variety of interactions leading to new kinds of knowledge artifacts. In keeping with the original vision of the internet, I imagine a future where any person can, with one friend or ten thousand strangers, have productive and meaningful discourse online.

REFERENCES

[1] **Amy X. Zhang**, Lea Verou, and David Karger. Wikum: Bridging discussion forums and wikis using recursive summarization. In Proceedings of CSCW 2017: *ACM Conference on Computer Supported Cooperative Work and Social Computing*, Portland, Oregon, 2017.

[2] Jane Im, **Amy X. Zhang**, Chris Schilling, and David Karger. Deliberation and resolution on Wikipedia: A case study of requests for comments. In Proceedings of CSCW 2018*: ACM Conference on Computer Supported Cooperative Work and Social Computing*, New York, NY, 2018.

[3] **Amy X. Zhang** and Justin Cranshaw. Making sense of group chat through collaborative tagging and summarization. In Proceedings of CSCW 2018*: ACM Conference on Computer Supported Cooperative Work and Social Computing*, New York, NY, 2018. **Best Paper Award**.

[4] Sacha Zyto, David Karger, Mark Ackerman, and Sanjoy Mahajan. Successful classroom deployment of a social document annotation system. In Proceedings of CHI 2012*: ACM Conference on Human Factors in Computing Systems*, 2012.

[5] Soya Park, **Amy X. Zhang**, and David Karger. Post-literate programming: Linking discussion and code in software development teams. In Proceedings of UIST 2018*: ACM Conference on User Interface Systems and Technology Symposium: Poster Publication*, 2018.

[6] **Amy X. Zhang**, Joshua Blum, and David Karger. Opportunities and challenges around a tool for social and public web activity tracking. In Proceedings of CSCW 2016*: ACM Conference on Computer-Supported Cooperative Work & Social Computing*, San Francisco, CA, 2016.

[7] **Amy X. Zhang**, Michele Igo, Marc Facciotti, and David Karger. Using student annotated hashtags and emojis to collect nuanced affective states. In Proceedings of L@S 2017*: ACM Conference on Learning at Scale: Poster Paper*, 2017.

[8] **Amy X. Zhang**, Mark S. Ackerman, and David Karger. Mailing lists: Why are they still here, what's wrong with them, and how can we fix them? In Proceedings of CHI 2015*: ACM Conference on Human Factors in Computing Systems*, Seoul, Korea, 2015.

[9] Soya Park, **Amy X. Zhang**, David Karger, and Luke Murray. Opportunities for automating email processing: A need-finding study. In Proceedings of CHI 2019*: ACM Conference on Human Factors in Computing Systems*, Glasgow, UK, 2019.

[10] Kaitlin Mahar, **Amy X. Zhang**, and David Karger. Squadbox: A tool to combat email harassment using friend-sourced moderation. In Proceedings of CHI 2018*: ACM Conference on Human Factors in Computing Systems*, Montreal, Canada, 2018.

[11] **Amy X. Zhang**, Aditya Ranganathan, Sarah Emlen Metz, Scott Appling, Connie Moon Sehat, Norman Gilmore, Nick B. Adams, Emmanuel Vincent, Jennifer Lee, Martin Robbins, Ed Bice, Sandro Hawke, David Karger, and An Xiao Mina. A structured response to misinformation: Defining and annotating credibility indicators in news articles. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, Lyon, France, 2018.

[12] **Amy X. Zhang**, Bryan Culbertson, and Praveen Paritosh. Characterizing online discussion using coarse discourse sequences. In Proceedings of ICWSM 2017*: AAAI Conference on Weblogs and Social Media*, Montreal, Canada, 2017.

[13] **Amy X. Zhang** and Scott Counts. Modeling ideology and predicting policy change with social media: Case of same-sex marriage. In Proceedings of CHI 2015*: ACM Conference on Human Factors in Computing Systems*, Seoul, Korea, 2015. **Best Paper Honorable Mention**.

[14] **Amy X. Zhang** and Scott Counts. Gender and ideology in the spread of anti-abortion policy. In Proceedings of CHI 2016*: ACM Conference on Human Factors in Computing Systems*, San Jose, CA, 2016.

[15] Nicholas Diakopoulos, **Amy X. Zhang**, Dag Elgesem, and Andrew Salway. Identifying and analyzing moral evaluation frames in climate change blog discourse. In Proceedings of ICWSM 2014*: AAAI Conference on Weblogs and Social Media*, Ann Arbor, MI, 2014.

[16] **Amy X. Zhang**, Jilin Chen, Wei Chai, Jinjun Xu, Lichan Hong, and Ed Chi. Evaluation and refinement of clustered search results with the crowd. *ACM Trans. Interact. Intell. Syst.*, 8(2):14:1–14:28, June 2018.