Paper ID #51581

From Curiosity to Capability: Expanding Pathways into Data and AI through ExLENT

Valentina Kuskova, University of Notre Dame

Dr. Valentina (Valya) Kuskova is a Professor and Associate Director at the Lucy Family Institute for Data & Society at the University of Notre Dame. She holds a Ph.D. in Decision Sciences, an MBA, and master's degrees in Applied Statistics and Business, along with a B.S. in Psychology. Professor Kuskova is dedicated to exploring how artificial intelligence can be leveraged to address complex societal challenges, including systemic inequities in access to healthcare education and careers in underserved communities, the socioeconomic impact of large corporations and the measurement of corporate social responsibility, and the systemic network factors driving poverty in disadvantaged regions. Her innovative work bridges advanced analytics, network analysis, and social sciences, driving impactful solutions across diverse domains such as social systems, corporate strategy, and public policy. Kuskova employs advanced AI techniques, including graph neural networks (GNNs), large language models (LLMs), natural language processing (NLP), neural network modeling, and AI-driven forecasting, to analyze and enhance complex problems, enabling diverse applications across various fields. She was the founding director of a network analytics institute and established and supervised two master's programs in Advanced Analytics and Network Analysis. Dr. Kuskova has extensive experience collaborating with corporate clients, including Coca-Cola, McDonald's, Volkswagen, the Association of European Businesses, and many others, serving as principal investigator, co-principal investigator, or advanced analytics methodologist. She is passionate about applying scientific research to create meaningful impact across industry and society. At the Lucy Family Institute for Data & Society, Professor Kuskova serves as a Co-Principal Investigator (Co-PI) of several educational initiatives. She plays a key role in shaping a holistic educational experience that bridges theoretical rigor with practical, real-world engagement. Professor Kuskova co-supervises interdisciplinary programs such as the Lucy Internship Program, and is the faculty of the new Responsible and Ethical AI Course developed in partnership with the National Education Opportunity Network, as well as a range of targeted workshops. Dr. Kuskova serves as the lead Principal Investigator (PI) of the NSF ExLENT Explorations Track project, Bridging to AI and Data Science Careers through Experiential Learning and Training. This project advances the national interest by designing and implementing a transformative experiential learning program in data science and artificial intelligence for undergraduate students and adult learners. In addition, Professor Kuskova is a Co-Principal Investigator (Co-PI) of the Ascension Foundation–funded #GOALS Evaluation Study project, which aims to evaluate how #GOALS (Go Out and Love Science) events, organized by Ascension Foundation, inspire middle-school students from underserved communities to pursue careers in healthcare. Collectively, these initiatives advance the Institute's mission to prepare students to address complex societal challenges through data-driven inquiry and innovation.

Dmitry Zaytsev, University of Notre Dame
Richard P. Johnson, University of Notre Dame

Rick Johnson serves as an associate professor of the practice and managing director of the Applied Analytics and Emerging Technology Lab (AETL) as part of the Lucy Family Institute for Data & Society at the University of Notre Dame. AETL collaborates with iNDustry Labs to help accelerate economic and innovation impact in the South-Bend Elkhart region, achieving the goals of the LIFT Network. He co-leads a data and AI focused internship program for undergraduate students at the University of Notre Dame. His research interests include data and AI platforms and systems, human computer interaction, and informatics.

Previously within the Hesburgh Libraries at Notre Dame, Johnson has combined his research, industry, and technology experience as the Director of the Strategic Innovation Lab which analyzed, prototyped, and vetted mid-to-long-term strategic opportunities for the Hesburgh Libraries. Rick also spearheaded the implementation of the University of Notre Dame's institutional repository, CurateND, contributed to several data preservation research grants, and served as a Visiting Program Officer for the Association of Research Libraries for the SHARE project.

Prior to joining the university in 2008, Rick worked as a software developer in the banking industry in Seattle for seven years, and spent two years in software sales and consulting in Chicago and Tokyo. He holds two bachelor's degrees from the University of Notre Dame in Computer Science and Japanese, and a Masters of Library and Information Science from the University of Wisconsin-Milwaukee.

Prof. Nitesh Chawla
Sugana Chawla, University of Notre Dame

Sugana Chawla is an Associate Professor of the Practice at the Lucy Family Institute for Data & Society at the University of Notre Dame and serves as the Data Science Education Program Director for iTREDS. Her work focuses on designing and leading innovative educational programs that integrate data science, artificial intelligence, and real-world applications. She is the Principal Investigator of the #GOALS project, where she leads research on using AI-driven platforms to explore student engagement and career pathways in STEM and healthcare. Her research has been presented at leading venues such as AIED and centers on advancing responsible AI in education. Sugana actively contributes to interdisciplinary curriculum development and collaborates with university and external partners to expand access to AI and data science education. She also serves as Education Chair for the RISE AI Conference, supporting broader engagement in AI-focused learning initiatives. She holds a an MA in Education and a BS in Environmental Science and Policy, and her work reflects a strong commitment to student learning, innovation, and impact.

From Curiosity to Capability: Expanding Pathways into Data and AI through NSF ExLENT (WIP)

Valentina V. Kuskova[1], Dmitry Zaytsev[1], Richard Patrick Johnson[1], Nitesh Chawla[2], Sugana Chawla[1]

Background: the NSF ExLENT Program

The National Science Foundation's Experiential Learning for Emerging and Novel Technologies (ExLENT) program is a strategic initiative designed to broaden pathways into careers in emerging technology fields by expanding access to high-quality, hands-on learning opportunities. The program addresses the dynamic nature of technologies such as artificial intelligence, biotechnology, advanced manufacturing, quantum information science, and microelectronics by supporting experiential learning that builds practical skills and workforce readiness for individuals across a wide spectrum of educational and professional backgrounds, including those with nontraditional STEM trajectories and adult learners seeking to upskill or reskill.

ExLENT emphasizes cross-sector partnerships among industry, government, educational institutions, non-profits, and workforce development organizations to develop scalable models that align learner preparation with regional economic needs. Through multiple tracks tailored to diverse learner stages, from foundational exploration to career pivots and deepening existing competencies, ExLENT seeks to lower barriers to participation, foster mentorship and community, and contribute evidence-based insights on effective practices for preparing a competitive and inclusive emerging technology workforce that strengthens U.S. innovation capacity and competitiveness.

Explorations: Bridging to AI and Data Science Careers

The overarching goal of this funded proposal is to develop a transformative experiential education program that brings together learners from different disciplinary backgrounds and life stages to explore data science collaboratively. This program - **Data Crossings: Bridging to Data Science Careers through Experiential Learning and Training** - prepares participants for domain-driven, problem-oriented, higher-order thinking needed to address complex societal and industry challenges, including healthcare, transportation, housing, and AI-enabled decision-making. The year-long mixed-cohort program engages non-STEM undergraduate students and adult learners with domain expertise who seek to acquire data science skills, while simultaneously partnering with community organizations and small- to medium-sized enterprises (SMEs) that often lack access to advanced data and analytics capacity [1].

Sustaining U.S. leadership in innovation requires educational models that cultivate translational skills and societal impact [2]. Data Crossings advances this goal through community- and industry-embedded experiential learning, including capstone projects and in-situ engagements, while intentionally developing leadership and STEM democratization competencies. Without such approaches, existing digital divides, particularly between SMEs, nonprofits, and large

---

[1] Lucy Family Institute for Data & Society, University of Notre Dame
[2] Department of Computer Science and Engineering, University of Notre Dame

corporations, will continue to widen, especially in the adoption of advanced data and analytics technologies [1].

The need for this program is driven by the accelerating pace of scientific and technological change. While disciplinary advances deepen understanding within fields [3], major societal challenges, such as climate change, poverty, and social inequality, require multidisciplinary collaboration among educators, industry partners, and practitioners. Experiential learning plays a critical role in bridging theory and practice, enabling learners to develop applied problem-solving and critical thinking skills through project-based research, internships, and simulations.

Simultaneously, the rapid growth of data availability and AI-driven analytical methods has transformed knowledge production across the social, natural, life, and physical sciences [4]. Domain-driven and data-driven approaches are increasingly fused, making computational, data, and AI competencies essential across disciplines. Fields such as bioinformatics illustrate how data-intensive methods have become mainstream within domain sciences rather than peripheral collaborations [5].

Demand for a STEM-literate workforce continues to grow. National policy emphasizes STEM education as essential to U.S. competitiveness and security [6], particularly as technological change outpaces workforce skill development [7]-[8]. Over several decades, interdisciplinary education models have emerged across engineering, medicine, information systems, psychology, and digital transformation [9], alongside initiatives to expand science and data literacy among non-STEM majors [10] and to upskill the existing workforce [11]. Despite these efforts, demand for data science literacy continues to exceed supply.

This mismatch is particularly evident in Indiana, where job postings requiring data literacy have more than doubled in five years [12]-[13], with data science among the fastest-growing sectors in 2024 [14]. Growth in digital transformation, accelerated by the COVID-19 pandemic, has further intensified workforce demand [15]. Contributing factors include low technology investment per worker [16], persistent skills mismatches [16]-[17], and rapid adoption of Industry 4.0 technologies requiring substantial upskilling of existing employees [18].

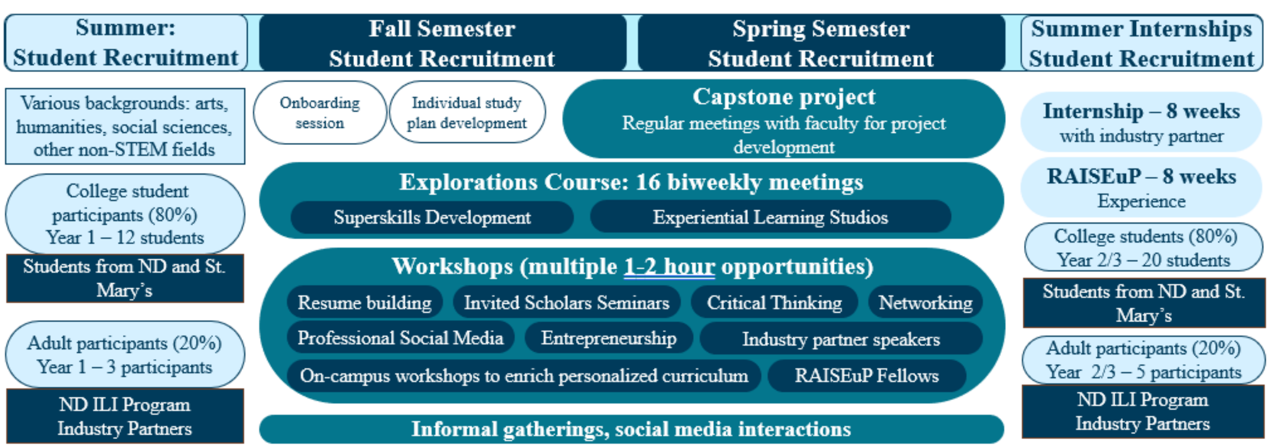

Figure 1. Overview of the Data Crossings Program Structure

Beyond workforce shortages, employers report significant gaps in interdisciplinary and applied skills [19]-[20]. Traditional academic programs often remain siloed and limited in accessibility, while adult learners face barriers to effective reskilling, including low retention in distance learning environments [21]-[22]. These conditions underscore the need for experiential, mixed-cohort educational models that integrate technical, analytical, and communication skills [23].

Aligned with NSF Strategic Goals 1 and 3 [24], Data Crossings addresses these challenges by extending the mixed-cohort model [8] to include college students and adult participants - learners across age, career stage, and disciplinary background - integrating domain-informed applications, and delivering targeted training in AI and data science. Through year-long instruction followed by immersive summer experiential learning, the program fosters STEM literacy, collaborative problem-solving, and the "superskills" [25] required for sustained societal and economic impact. The program's essence is presented in Figure 1.

Lessons from First Year Applicant Cohort

Methods. The first-year call for participants attracted over 200 applications. Applicants were asked to respond to a prompt describing why they believed they were a strong fit for the program. These narrative responses constituted the primary dataset for analysis, and each complete applicant narrative was treated as a single unit of analysis.

We conducted deductive-inductive thematic content analysis [26]-[27], augmented by NLP-assisted pattern extraction [28]-[29]. The analytic process proceeded in four stages. First, we performed iterative close reading of all narratives to identify recurring concepts, value statements, and motivational patterns. This phase allowed for inductive theme emergence while also attending to deductive categories derived from the program's design goals (e.g., interdisciplinary identity, experiential learning orientation, ethical AI framing).

Second, we used NLP-assisted techniques to support theme refinement and consistency. We examined high-frequency unigrams and multi-word expressions (bigrams and trigrams) to identify recurrent lexical clusters across narratives. We also analyzed co-occurrence patterns among key terms (e.g., "AI" with "ethics," "policy," or "social good"; "data" with "translation," "communication," or "impact") to detect thematic relationships not immediately apparent through manual reading alone. In addition, we reviewed semantic similarity patterns across documents to identify conceptually related narratives that used different vocabulary to express similar motivations (e.g., "bridging disciplines," "connecting worlds," "translating knowledge").

Third, emergent and theory-informed themes were formalized into a structured coding framework. Each narrative was then coded for the presence or absence of each theme using binary document-level coding. Themes were not mutually exclusive; a single narrative could reflect multiple motivations simultaneously. Finally, descriptive frequencies were calculated as the proportion of applicant narratives in which a given theme was present. Reported percentages therefore reflect document-level thematic presence rather than word frequency or probabilistic topic proportions.

Throughout the process, NLP outputs were treated as heuristic aids rather than deterministic classifications; final theme definitions and coding decisions remained interpretive and document-level. This approach aligns with computational grounded theory and text-as-data methodologies that integrate algorithmic pattern recognition with human analytic judgment [28]-[29].

Results. Across approximately two hundred narrative applicant narratives, several clear patterns emerged. The dominant pattern is a strong, self-aware demand for foundational AI and data science literacy (T1; ~89%), frequently framed not as a pathway to becoming engineers, but as a mechanism for translating domain expertise into socially meaningful impact. Rather than emphasizing computational depth, applicants emphasized ethical application, interdisciplinary integration, and the ability to interpret and communicate data responsibly.

This technical upskilling is rarely presented as an end in itself. Instead, applicants consistently articulate a public-interest rationale for learning data methods (T2; ~73%), emphasizing social good, community benefit, and more equitable decision-making. A complementary and highly prevalent theme is interdisciplinary identity (T3; ~69%): applicants position themselves as "bridge" actors who want to translate between qualitative domains (policy, humanities, design, nonprofit leadership) and technical tools, suggesting that the program is perceived as an enabling infrastructure for cross-domain collaboration rather than a conventional STEM pipeline.

Applicants also strongly prefer experiential learning modalities (T4; ~64%), frequently citing hands-on studios, internships, and capstone projects as the mechanism by which technical ideas become usable in real settings. Importantly, ethics and responsible use appear as a core expectation rather than an add-on (T5; ~60%), with references to governance, bias, accountability, and harms ranging from misinformation to unequal impacts. Substantively, the most common application domains include public policy and governance (T6; ~49%), communication and translation of insights to stakeholders (T7; ~40%), and career pivot or career acceleration motivations (T8; ~40%). A sizable subset explicitly link participation to research trajectories, including senior theses and graduate study (T9; ~36%).

Finally, several specialized but meaningful clusters indicate where "data crossings" is most salient: digital humanities and text-as-data (T10; ~24%), NLP and language-learning quality (T11; ~20%), business/marketing analytics (T12; ~20%), and design/architecture applications (T13; ~16%). Smaller yet policy-relevant clusters highlight cybersecurity and national security (T14; ~13%), climate and development (T15; ~13%), and epistemic risk concerns such as deepfakes and hallucination detection (T16; ~9%). Taken together, these thematic distributions characterize a cohort motivated by applied, ethically grounded, interdisciplinary learning; seeking not only technical competence, but also the capacity to translate data tools into credible decisions and social impact.

Beyond the distribution of themes, the narratives reveal several interpretive patterns that are central to understanding how applicants conceptualize both AI and the role of this program. First, applicants consistently position themselves not as aspiring engineers or technical specialists, but as *boundary-spanners* operating between domains. Many describe prior training in policy, humanities, business, design, education, or public service, and frame their motivation as the desire to “bridge,” “translate,” or “connect” these domains with emerging data tools. The language of integration appears far more frequently than the language of specialization. Rather than seeking immersion in advanced computational theory, applicants seek the capacity to embed AI and data methods within existing professional identities. This suggests that the demand is not for conversion into STEM disciplines, but for augmentation of domain expertise with technical fluency.

Second, ethics emerges not as an abstract philosophical concern but as an operational one. Applicants reference algorithmic bias, model training data, governance structures, regulatory oversight, misinformation, hallucination, and deployment risks. Ethics is frequently discussed in relation to concrete mechanisms: how models are trained, how data are selected, how outputs influence decisions, and how systems affect vulnerable populations. This framing indicates that participants do not perceive ethical literacy as separate from technical competence; rather, they understand responsible AI as requiring insight into model behavior and data provenance. In this sense, ethical awareness is tightly coupled with epistemic competence.

Third, there is a clear preference for experiential translation over computational depth. While foundational technical literacy is strongly desired, few narratives call for advanced mathematics or programming mastery. Instead, applicants emphasize “hands-on,” “real-world,” “internship,” and “capstone” experiences as the locus of meaningful learning. Data competence is framed as the ability to interpret outputs, communicate findings, and apply tools in context. This suggests that participants value applied reasoning, stakeholder communication, and decision-oriented analysis over purely technical optimization. The emphasis on translation reinforces the boundary-spanning identity noted above.

Finally, applicants frequently describe the program not simply as skills training, but as a formative transition point in their professional trajectory. Many articulate the program as a bridge between academic identity and workforce relevance, or between qualitative expertise and quantitative capability. In this way, Data Crossings is perceived less as a discrete training intervention and more as a career identity inflection point. Participants appear to view the program as enabling them to reposition themselves within a rapidly evolving technological landscape, without abandoning their core disciplinary commitments.

These findings suggest that the program meets a distinct and under-served demand: learners who do not seek to become engineers, but who require sufficient technical literacy to responsibly integrate AI into governance, policy, design, research, and community contexts. This interpretive layer clarifies how the applicant cohort aligns with the program’s mixed-cohort, experiential, and ethically grounded design, and highlights broader workforce implications for interdisciplinary AI education. While the first cohort has just started the program, from the applicant perception of the program, Data Crossings will play a critical role in enabling non-STEM learners to transform ethical awareness and interdisciplinary insight into practical, workforce-relevant data and AI competencies that advance societal benefit.

Acknowledgement. This material is based upon work supported by the National Science Foundation under Grant No. 2526307.

References

[1] OECD, "OECD Studies on SMEs and Entrepreneurship The Digital Transformation of SMEs," 2021. Available: https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/02/the-digital-transformation-of-smes_ec3163f5/bdb9256a-en.pdf

[2] K. Ramsey, and B. Baethe, "The Keys to Future STEM Careers: Basic Skills, Critical Thinking, and Ethics." *Delta Kappa Gamma Bulletin*, vol. 80, no 1., p. 26.

[3] S. Gehlert, K. L. Hall, and L. A. Palinkas, "Preparing Our Next-Generation Scientific Workforce to Address the Grand Challenges for Social Work," *Journal of the Society for Social Work and Research*, vol. 8, no. 1, pp. 119–136, Mar. 2017, doi: https://doi.org/10.1086/690659.

[4] C. Challoumis, "*The Imperative of Skill Development in an AI Revolution*." XIX International Scientific Conference. London. Great Britain, pp. 132-168.

[5] J. B. Hagen, "The origins of bioinformatics," *Nature Reviews Genetics*, vol. 1, no. 3, pp. 231–236, Dec. 2000, doi: https://doi.org/10.1038/35042090.

[6] The White House, "National Security Strategy," The White House, Oct. 2022. Available: https://www.whitehouse.gov/wp-content/uploads/2022/10/Biden-Harris-Administrations-National-Security-Strategy-10.2022.pdf

[7] "Ambition Four: Create an ecosystem for collective learning and sharing to ensure opportunities for excellence for all: Teacher Professional Learning," *OECD*, 2026. https://www.oecd.org/en/publications/teacher-professional-learning_0cceeddf-en/full-report/ambition-four-create-an-ecosystem-for-collective-learning-and-sharing-to-ensure-opportunities-for-excellence-for-all_7967922d.html (accessed Feb. 16, 2026).

[8] A. Schleicher, "The case for 21st-century learning - OECD," *www.oecd.org*, 2010. https://www.oecd.org/general/thecasefor21st-centurylearning.htm

[9] M. Brady, "SingleDiscipline Schooling," *The Phi Delta Kappan*, vol. 74, no. 6, pp. 438–443, 1993, doi: https://doi.org/10.2307/20404907.

[10] "STEM for non-STEM Majors: Enhancing Science Literacy in Large Classes," *Journal of College Science Teaching*, 2026, doi: https://doi.org/10.2505//4//jcst13_042_06_20.

[11] G. Li, C. Yuan, S. Kamarthi, M. Moghaddam, and X. Jin, "Data science skills and domain knowledge requirements in the manufacturing industry: A gap analysis," *Journal of Manufacturing Systems*, vol. 60, pp. 692–706, Jul. 2021, Available: https://www.sciencedirect.com/science/article/pii/S0278612521001448

[12] "Seismic Shifts in the Talent Landscape: 2023 TechPoint Indiana Tech Workforce Report," 2023. Accessed: Feb. 16, 2026. [Online]. Available: https://techpoint.org/wp-content/uploads/2023/01/2023-TechPoint-Tech-Workforce-Report.pdf

[14] "Mission41K Update Tech Workforce Data Update." Accessed: Feb. 16, 2026. [Online]. Available: https://techpoint.org/wp-content/uploads/2024/11/mission41K-update-report-November-2024.pdf

[15] https://techpoint.org/

[16] J. Merisotis, "Op-ed: Manufacturing tech is leaving Indiana behind. Here's what we can do about it.," *The Indianapolis Star*, May 12, 2021. https://www.indystar.com/story/opinion/2021/05/12/lumina-foundation-ceo-jamie-merisotis-indiana-manufacturing-tech-jobs/4980899001/ (accessed Feb. 16, 2026).

[17] P. C. Verhoef *et al.*, "Digital transformation: a Multidisciplinary Reflection and Research Agenda," *Journal of Business Research*, vol. 122, no. 122, pp. 889–901, Jan. 2021, doi: https://doi.org/10.1016/j.jbusres.2019.09.022.
[18] N. Dmitrovich, "Tech is Growing Fast in Manufacturing," *Building Indiana Business*, Jun. 04, 2022. https://buildingindiana.com/tech-is-growing-fast-in-manufacturing/ (accessed Feb. 16, 2026).
[19] "Career Pathways and Competency Maps." Available: https://techpoint.org/wp-content/uploads/2023/02/Techpoint-Tech-Pathways-Competency-Maps-12.9.22.pdf
[20] E. S. Jagannathan. "Reimagining Digital Learning for Sustainable Development: How Upskilling, Data Analytics, and Educational Technologies Close the Skills Gap," *Sadil.ws*, 2021, doi: https://doi.org/13:%20978-0367540180.
[21] K. Komp-Leukkunen *et al.*, "Older Workers in Digitalizing Workplaces: A Systematic Literature Review," *The Journal of Aging and Social Change*, vol. 12, no. 2, pp. 37–59, 2022, doi: https://doi.org/10.18848/2576-5310/cgp/v12i02/37-59.
[22] M. Kara, F. Erdoğdu, M. Kokoç, and K. Cagiltay, "Challenges Faced by Adult Learners in Online Distance Education: a Literature Review," *Open Praxis*, vol. 11, no. 1, pp. 5–22, Mar. 2019, doi: https://doi.org/10.5944/openpraxis.11.1.929.
[23] P. S. Muljana and T. Luo, "Factors Contributing to Student Retention in Online Learning and Recommended Strategies for Improvement: A Systematic Literature Review," *Journal of Information Technology Education: Research*, vol. 18, pp. 019–057, 2019, doi: https://doi.org/10.28945/4182.
[24] "NSF Strategic Plan | NSF - National Science Foundation," *www.nsf.gov*. https://www.nsf.gov/news/special_reports/strategic_plan/
[25] "Harnessing the Data Revolution (HDR): Data Science Corps (DSC)," *NSF - U.S. National Science Foundation*, 2018. https://www.nsf.gov/pubs/2019/nsf19518/nsf19518.htm (accessed Feb. 16, 2026).
[26] V. Braun and V. Clarke, "Using Thematic Analysis in Psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006, doi: https://doi.org/10.1191/1478088706qp063oa.
[27] M. Forbes, "Thematic analysis: A practical guide," *Evaluation Journal of Australasia*, vol. 22, no. 2, p. 1035719X2110582, Dec. 2021, doi: https://doi.org/10.1177/1035719x211058251.
[28] J. Grimmer and B. M. Stewart, "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," *Political Analysis*, vol. 21, no. 03, pp. 267–297, 2013.
[29] L. K. Nelson, "Computational Grounded Theory," *Sociological Methods & Research*, vol. 49, no. 1, p. 004912411772970, Nov. 2017, doi: https://doi.org/10.1177/0049124117729703.