Promoting Ethical Awareness in Communication Analysis: Investigating Potentials and Limits of Visual Analytics for Intelligence Applications

Maximilian T. Fischer
max.fischer@uni-konstanz.de
University of Konstanz
Konstanz, Germany

Simon David Hirsbrunner
simon.hirsbrunner@uni-tuebingen.de
IZEW, University of Tübingen
Tübingen, Germany

Wolfgang Jentner
wolfgang.jentner@uni-konstanz.de
University of Konstanz
Konstanz, Germany

Matthias Miller
matthias.miller@uni-konstanz.de
University of Konstanz
Konstanz, Germany

Daniel A. Keim
keim@uni-konstanz.de
University of Konstanz
Konstanz, Germany

Paula Helm
paula.helm@uni-tuebingen.de
IZEW, University of Tübingen
Tübingen, Germany

ABSTRACT

Digital systems for analyzing human communication data have become prevalent in recent years. This may be related to the increasing abundance of data that can be harnessed but can hardly be managed manually. Intelligence analysis of communications data in investigative journalism, criminal intelligence, and law present particularly interesting cases, as they must take into account the often highly sensitive properties of the underlying operations and data. At the same time, these are areas where increasingly automated, sophisticated approaches and tailored systems can be particularly useful and relevant, especially in terms of Big Data manageability. However, by the shifting of responsibilities, this also poses dangers . In addition to privacy concerns, these dangers relate to uncertain or poor data quality, leading to discrimination and potentially misleading insights. Other problems relate to a lack of transparency and traceability, making it difficult to accurately identify problems and determine appropriate remedial strategies.

Visual analytics combines machine learning methods with interactive visual interfaces to enable human sense- and decision-making. This technique can be key for designing and operating meaningful interactive communication analysis systems that consider these ethical challenges. In this interdisciplinary work, a joint endeavor of computer scientists, ethicists, and scholars in Science & Technology Studies, we investigate and evaluate opportunities and risks involved in using Visual analytics approaches for communication analysis in intelligence applications in particular. We introduce, at first, the common technological systems used in communication analysis, with a special focus on intelligence analysis in criminal investigations, further discussing the domain-specific ethical implications, tensions, and risks involved. We then make the case of how tailored Visual Analytics approaches may reduce and mitigate the described problems, both theoretically and through practical examples. Offering interactive analysis capabilities and what-if explorations while facilitating guidance, provenance generation, and bias awareness (through nudges, for example) can improve analysts' understanding of their data, increasing trustworthiness, accountability, and generating knowledge. We show that finding Visual Analytics design solutions for ethical issues is not a mere optimization task with an ideal final solution. Design solutions for specific ethical problems (e.g., privacy) often trigger new ethical issues (e.g., accountability) in other areas. Balancing out and negotiating these trade-offs has, as we argue, to be an integral aspect of the system design process from the outset. Finally, our work identifies existing gaps and highlights research opportunities, further describing how our results can be transferred to other domains. With this contribution, we aim at informing more ethically-aware approaches to communication analysis in intelligence operations.

CCS CONCEPTS

• **Computing methodologies** → **Visual analytics**; *Natural language processing*; • **Social and professional topics** → **Computing / technology policy**.

KEYWORDS

Communication Analysis, Visual Analytics, Intelligence Analysis, Ethic Awareness, Science & Technology Studies, Critical Algorithm Studies, Critical Data Studies, Machine Learning, Interdisciplinary Research

ACM Reference Format:
Maximilian T. Fischer, Simon David Hirsbrunner, Wolfgang Jentner, Matthias Miller, Daniel A. Keim, and Paula Helm. 2022. Promoting Ethical Awareness in Communication Analysis: Investigating Potentials and Limits of Visual Analytics for Intelligence Applications. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3531146.3533151

1 INTRODUCTION

In recent years, the share of human communication transitioning to digital forms has increased significantly, while the advances in big data analysis offer new opportunities concerning the amount of

FAccT '22, June 21–24, 2022, Seoul, Republic of Korea

meaningful information and knowledge that can be harnessed from this data. Consequently, advanced digital systems for analyzing such communication have simultaneously become increasingly prevalent, especially due to the difficulty in managing such large and diverse information either manually or with only rudimentary analysis capabilities. Novel approaches allow for more effective handling, performing sophisticated big data analyses using methods such as social network analysis [92], meta-data screening [22], pattern matching [44], or natural language processing [24], often assisted through machine learning [33].

One of the most prevailing domains for such systems is the analysis of communications data for intelligence purposes, namely in criminal investigations, lawsuits, matters of national and international security and in investigative journalism. Specialized systems are used, for example, by the National Security Agency (NSA) as part of global spying operations [77] or law enforcement against organized crime [23], by lawyers for analyzing case-relevant documents [2], but also by journalists working on [103] large data leaks such as the Panama Papers. During these operations, large amounts of communication data, like e-mails, chats, posts, or calls are collected, along with associated documents (e.g., attachments) and meta-data like timestamps, locations, and contact networks. In our research context, these domains present a particularly interesting case, as they should consider the often highly sensitive and private character of the underlying operations and data with particular caution. There is no doubt that untargeted mass collection of communication in the name of national security is privacy-invasive and thus highly controversial [71]. However, many ethical challenges remain even relevant for morally more accepted cases like specifically targeted analysis of confiscated organized crime equipment or even practices considered essential to democratic culture and particularly valuable, like data journalism.

For example, privacy issues relating to the separation of irrelevant data, its secure handling, analysis, and deletion have to be considered [6, 10]. Poor data quality, unreliable methods, or biased algorithms may lead to misleading insights [3], bearing the risk of overlooking critical information, or worse, contribute to discriminatory practices against people of colour [16], cement existing social inequalities [26], and may even result in false accusations [81]. Moreover, a lack of transparency can make it impossible to defend oneself against such accusations [94] when the systems supporting (or making) such decisions are considered reliable, but actually do not (consistently) provide complete chains of evidence [61]. Unfortunately, being focused on efficiency and quick results, not all actors consider the arising ethical challenges in this field with a long-term perspective in mind, and even those who try, may be limited by their technical approach and implementation difficulties, for example, through black-box machine learning models or an incomplete understanding of the considerations involved in deriving the result [104].

The concept of accountability in computer science [78] stresses the need to handle and answer to the harms and risks that can be caused by technology. While this concept has been around for decades already, concrete ways of handling accountability in digital analytical systems have remained vague. Yet, progress has been made in interpretability of machine learning [76, 88]. Whereas the coming into effect of the EU General Data Protection Regulation (GDPR) lead to more awareness on this topic and its implementation [49], its legal effects on the fields under consideration are limited due to exception clauses in article 2 (law enforcement) and article 85 (journalism) [30]. Even in those areas, however, the need to consider these topics carefully is increasingly prevailing.

Working together closely with criminal investigators from various institutions, we know of a growing awareness of these difficulties, also on the part of the analysts themselves. Concerns about the trustworthiness of their analytics systems and the ethical considerations involved have been expressed. This concern is also reflected in digital communication analytics in general [101], where the need for more detailed analysis was identified. To date, ethical concerns related to automated communications analysis have been described mainly from either a strictly sociological and/or ethical perspective [46] or in the context of technical capabilities [34, 36, 101]. Much less work, by contrast, addresses the complex techno-ethical tensions and dilemmas that arise in the messy gray areas of socio-technical feasibility, given the limits and consequences induced by the alleged solutions. Given the lack of overarching work, in this paper we examine ethical considerations in communications analysis for intelligence applications in more detail and propose possible mitigation techniques, which we discuss critically with regard to ethical concerns. Unlike most previous research, we thus bridge ethical considerations, sociological science & technology studies, and a computer science perspective.

We study concrete design approaches and solutions and by analyzing the interfacing problem from an interdisciplinary perspective, we can critically reflect on the opportunities and challenges involved [66]. We argue that designing solutions is not a mere optimization task, but balancing out and negotiating the trade-offs has to become an integral aspect of the design process at the very outset. We further claim that — in light of recent requirements for human oversight [28, 29] — an application design based on visual analytics principles is uniquely suited for such a task:

Visual analytics (VA) [57, 58] combines machine learning methods with interactive visual interfaces to enable human sense- and decision-making. VA features interactive visualization and data processing elements that analysts can engage in, shape and control, aiming at interactively combining human sense- and decision-making with computational power, leveraged through a frequent feedback loop. The overall iterative analysis process in VA [57] can be summarized as an explicit process model [89] for knowledge generation. It shows how a user is supported at every stage of the visual data analysis process from exploratory analysis over verification (confirmatory) of hypothesis to knowledge generation and sets this in relation to human sense-making, the information visualization pipeline [17] and the knowledge discovery in databases process [38] from data science. Through these aspects, VA can better handle ill-defined or open-ended tasks — which often occur in criminal intelligence, law, or investigative journalism — than fully-automated systems. Through human oversight, VA is well-suited for designing and operating meaningful interactive communication analysis systems that consider the ethical challenges outlined above and which we discuss in more detail in the following sections, ease their technical implementation, and allow for an interactive and iterative knowledge generation process [34].

With this work, we aim to promote ethical awareness of digital communication analysis by turning our focus to the interface, investigating the potentials and limitations of visual analysis for intelligence applications, contributing:

- A detailed **discussion** on the ethical frictions and tensions involved in intelligence applications, followed by a **scenario**-based stakeholder analysis of actors and their roles.
- A **critical reflection** of visual analytics design solutions fostering ethical awareness in communication analysis and the involved trade-offs as an integral part of the interface design process.

With this contribution, we aim to inform more ethically-aware approaches to communication analysis in intelligence operations using visual analytics principles.

2 ETHICAL CHALLENGES FOR COMMUNICATION ANALYSIS

AI ethics is an expanding field that both drives and responds to the various concerns associated with increasing datafication and automation. Novel technologies used in police intelligence are receiving particular attention in this context, as they are seen as extraordinarily problematic [94]. Their sometimes ill-advised use is being critiqued by both ethics scholars and civil society actors, as illustrated by the fierce backlash against so-called predictive policing technologies (PPTs) [3]. Several key challenges have been identified in the context of this growing debate. Below, we discuss those that are particularly relevant to communications analysis.

C1. Discriminatory bias – One of the most urgent challenges are biased algorithms reproducing existing stereotypes and aggravating discrimination of communities (e.g., women, people of color, transgender) [16, 20, 69, 79]. Problematic bias in machine learning may be introduced by human actors (e.g., suspicion based solely on ethnic belonging) or inserted through processes (e.g., replicating stereotypes represented in the training data), for example, in facial recognition [41] systems and pre-trained language models [63]. Correspondingly, fairness principles and metrics are developed to evaluate and mitigate statistical discrimination in AI models [14, 74]. These however may not be enough in societies that suffer from unfair conditions already and require equity before automation [43].

C2. Privacy – Privacy has been and remains to be a core challenge. Intelligence applications process enormous amounts of data collected from heterogeneous sources. This may include seized data like recorded phone calls, but also online messengers and social media. Due to quantity and heterogeneity, the majority of information is on unrelated third-parties. The storing and processing of such information therefore triggers important legal and ethical questions [75]. While law enforcement is exempt from many legal frameworks regulating data protection and privacy (most notably the GDPR [30]), this does not give police authorities a free pass. It instead puts more weight on the responsibility to safeguard against misuse of gathered information. Access to such sensible data, for example, has to be clearly defined, explained, and technically implemented in a robust way.

C3. Opacity – The black-box architecture of many deep-learning systems is another novel challenge [84]. Even if the systems are in principle explainable and/or interpretable by experts in the field, analysts are generally not such experts. It is therefore difficult for them to fathom how the systems they use generate their output [8]. In addition, there are the typical problems of public-private partnerships, such as when algorithms used in public domains fall under trade secret protection. This can lead to severe problems, as analysts, judges, the public in general, and those directly affected by the results have little opportunity to challenge the findings. However, these may well be biased or inaccurate. Sometimes with dramatic consequences. [69, 79, 81].

C4. Exaggerated Expectations – Another strand of criticism deals with the connotation that algorithmic recommendations cannot be disputed because they are mathematical truths. In contrast to the black-box discourse, the discourse that addresses this misunderstanding focuses on the problem of trustworthiness not primarily as a technical, but as an ideological one. Highlighting the opportunities of innovative technologies should not result in practices of "mathwashing" where software is used as a panacea against error-prone or malicious activities of humans [54]. This imaginary is triggered by the portrayal as charismatic machines [4, 5], with so-called predictive policing technologies being a prime example of this mismatch between promise and actual capability [46]. To counter imaginaries of the flawless machine, STS scholars draw attention to the various forms of subjectivity and intention that are woven into systems and result from collective decisions made by a variety of actors with diverse, potentially conflicting interests [61, 70]. Given the collective work and design decisions that underlie it, communications analysis of intelligence data is anything but neutral [42]. When people belief in the neutrality of tools that in fact serve particular needs and interests, they are deprived of their ability to question the results. This "erasure of doubt" is deeply problematic because it impedes reasoned trust grown from education and experience [7].

C5. Human-Machine-Configurations – With advanced automation, human machine configurations grow ever more complex [100]. The challenge here is to accomplish good integration and agree on an appropriate level of automation. The goal of automated analysis is to assist investigators by relieving them of scanning through irrelevant and mundane patterns so that they can focus on more useful activities [19]. Sometimes, however, analysts perceive their machine assistants as competition rather than help, and feel disregarded and displaced in their human experience [56]. The critical question is how much automation should be used, how far it should go, and how it should relate to human investigators' activities [99]. What operations need to be supported and guided, for example, by recommending alternative search terms or related individuals, and what agency should the analyst retain in interpreting the machine output? How much and what contextual information should be displayed, how much should be hidden for the sake of privacy and to what extent should interfaces be designed to encourage the user to consider ethical issues, e.g., through nudging? Smooth and well-structured collaboration is a much-discussed topic, as it is seen as a necessary safeguard against AI systems that may undermine human autonomy [27] with sometime detrimental effects.

C6. Accountability – The ethical issues mentioned so far subsequently trigger questions of accountability of the software, its designers, and users. Accountability refers to the willingness or

obligation to assume responsibility for actions and decisions of AI systems [1, 65]. In the context of AI, it has to be decided to what degree system users can and should be held accountable for consequential mistakes made if the software failed to meet basic standards of explainability and interpretability. It also needs to be specified what obligations the software provider has to safeguard against ethically-problematic decisions (racially-biased categorization of suspects) and usage (spying on third-parties).

3 SCENARIO ANALYSIS

As a prerequisite for the following discussion, we first provide a overview of communication analysis and the common technological systems, before presenting the PEGASUS research project as a case study. We then construct a hypothetical scenario, from which we derive a map of the stakeholders in conflict, forming the basis for our proposition for mitigation.

3.1 Digital Communication Analysis and Employed Technology

Digital communication analysis as a research field has no universally accepted definition, with different understandings in different domains. In this work, we follow the definition by Fischer et al. [34], considering it to encompass the computer-mediated [35] analysis of meaningful digital [91] information exchanges between humans [85]. The analysis relates not only to the actual content (text, audio, or video), but also encompasses accompanying metadata as well as communication network structures. Existing communication analysis approaches rarely consider these aspects holistically [35], but primarily focus on individual aspects: Most commonly, these are textual analysis through fuzzy search (and increasingly natural language processing (NLP) [72] methods) as well as social network analysis [92]. For example, in intelligence, one of the most commonly used systems [33] is IBM's i2 Analysts Notebook [51], which has a strong focus on network analysis and information management but has, so far, lacked advanced textual analysis capabilities. However, competing solutions such as Nuix [80], DataWalk [21] and Palantir Gotham [82] have been gaining ground [35]. Many are primarily large information management systems, using established algorithms (e.g., for centrality calculations in a network) and deterministic filters (e.g., keywords). Novel machine learning-based capabilities used for relevance scoring, person attribution, or facial matching are increasingly used in this context. The reliability of these models, however, the question of hidden bias, and the overall reproducibility (e.g., after updates), remain unclear.

In investigative journalism, tools like New/s/leak 2.0 [103], as used by *Der SPIEGEL*, use models trained on public data like Wikipedia for discovering named entities in textual data (e.g., persons or company names). Similarity, the industry-standard spacy [48] uses public corpora and increasingly open web information for model training. While this often results in increased accuracy, concerns about the reliability for less common languages or risks of manipulation (e.g., for datasets extracted from Wikipedia) remain valid.

3.2 The PEGASUS Research Project

For a case study on the requirements in intelligence, we specifically focus on the insights gathered through the work in the academic research project PEGASUS, funded by the Federal Ministry of Education and Research of Germany (BMBF). The project aims at improving big data analysis in the context of civil security, also considering the ethical challenges involved. The PEGASUS acronym — *not* to be confused with the unfortunately equally named PEGASUS spyware — stands for *Collection and analysis of heterogeneous Big Data by the police to fight organized crime structures*. Organized crime is a transnational and global form of crime, encompassing a broad spectrum of different areas, including human, drug, and arms trafficking, money laundering, smuggling operations, environmental, medical, cyber, and other white-collar crimes. According to Europol, in Europe alone, the number of criminal organizations under investigation is over 5000 (2017) [31], coming with high economic cost and a destabilizing effect on public security (through, for example, extortion, fraud, trafficking, or bodily harm). Organized crime can be characterized by its organized hierarchies (e.g., clans, mafia structures, shell companies) and sophisticated criminal acts using modern technology, and their ability to adapt quickly to changing circumstances [83] . For example, the COVID-19 pandemic has significantly affected organized crime, which was quick to adapt to new illegal avenues and modi operandi [31]. In conjunction, the seized data is increasing massively, overwhelming traditional (primarily manual) investigation methods. A significant share accumulates as intra- and inter-group communication and can be acquired, for example, when electronic devices are seized. However, the challenges faced are not unique to law enforcement; the goals are strikingly similar to tasks in fields such as investigative journalism and business intelligence, where information and the knowledge derived from it have become more important than natural resources [60]. Tackling the arising ethical issues is challenging because mitigation techniques incorporate numerous tensions and dilemmas that must be carefully weighed between the complex interplay of actively and passively involved stakeholders.

3.3 A Scenario in Police Intelligence Work

We construct a hypothetical scenario [18] of communication analysis within police intelligence work using current but non-visual analytics software that acts as a reference for our study of ethical challenges and emerging mitigation strategies. The scenario focuses on the challenges and practices of police officers and investigators as the main user community and points out other actors and stakeholders (highlighted as S Name) in an exemplary way.

S Martin is a police officer at the organized crime unit of the federal police. He currently investigates the selling of fake COVID-19 vaccination passports by an alleged criminal organization named *The Medics*. The Medics offer the counterfeit certificates to their S customers via the Telegram messenger. Unknown to Martin yet, S Chris , S Carlos , and S Eggert are Medics members, also communicating with their colleagues and suppliers via group Telegram channels while using pseudonyms, sometimes coded language, and images. In their free time, they also communicate with several friends, including their girlfriends, S Sarah and S Marta , who

are unaware of their business. Martin's police unit gathers much information about The Medics using traditional investigative methods. This information leads to the identification of the suspect, Chris, who seems to be a low-level member of The Medics. On one evening, Chris is found with blank vaccination certificates during a traffic stop. He is arrested, and his phone is seized by investigator Martin, who aims at using the information on the phone to track down the individuals pulling the strings. After calling judge S Robert to get a search warrant, which is granted, he then searches Chris's unprotected phone, finds the Telegram communication, and extracts it. He recalls that his superior, S Dr. D, asked him to try out the new AutoCommAnalyzer software, which was recently purchased from the multinational company AI-Tech Corp. The software purchase was part of a strong push by the government to digitally optimize work processes at the police forces. Martin looks at the training notes by the head developer S Molly, trying to remember how the machine learning-driven software — trained with texts by S Alf and S Bert — is supposed to direct him to the relevant communication. The software presents him with the most frequent contacts, with Sarah on top. He reads through this communication, as the software has flagged several words like package and hospital, discovering some explicit images but finding that the flags refer to a delivery package and a hospital stay for a broken ankle. In a second chat, the AI highlighted several currency amounts, and manually reading through it, it becomes clear that expensive *"stamps"* have been sold. Luckily for him, many addresses and names are also included in the chat messages. Searching for all chats that talk about stamp selling, he also finds one with Carlos, including his last name, and one from a person called Big E, which includes an address. Using the nationwide register, he finds a person named Carlos, who used to be a roommate with Chris, and only one person named Eggert is living at the found address. After completing his analysis, he finishes his report and submits it for the trial. During the court proceedings weeks later, Martin is questioned by judge S Muller on his findings. Ultimately Chris, Carlos, and Eggert admit to their guilt and are sentenced for document falsification. The intelligence gathered points to other alleged criminal networks and informs other running investigations.

3.4 Conflicts of Interest between Different Stakeholders

A stakeholder analysis based on the previous scenario helps to identify, map and describe the different actors and their roles involved in the scenario (see Figure 1), with the interdependent individuals having potentially conflicting interests. We propose four main groups of stakeholders: *civil society*, *governmental authorities*, *software provider*, and *data subjects*. This categorization has to be understood as a heuristic with possible overlaps and without claims of being exhaustive.

Data Subjects — Following the concept of data subjects defined by the GDPR [30], we consider the role of natural persons and their data ownership. Immediately apparent becomes the role of the **targets** (S Chris). In many cases, a target is unknown, but one has a list of **suspected targets** (S Carlos), indicating a different degree of certainty. One issue of communication analysis, however, is that communication is not strictly separated and touches on many other data subjects. These can be as of yet **unidentified** persons (e.g., the customers), but also **third parties** (e.g., S Sarah). With the use of machine learning, a fifth subgroup emerges, the **training data subjects** (S Alf and S Bert), whose data is leveraged as part of training the weights in neural networks. Further conflicts of interest arise between uninvolved third parties who usually (and rightly) do not want to be involved in a privacy-invasive investigation, which can also apply to (unwitting) training data subjects. A delicate issue are privacy considerations in the face of imminent suspicion: while target subjects clearly do not want to be investigated either, the reasons and justifications here differ substantially to those of uninvolved third-parties.

Governmental Authorities — In this specific type of intelligence analysis, the opposite of the data subjects are governmental authorities, with their investigating bodies. Here, because this applies to our setting, we assume a democratic political system that follows a separation of executive, legislative and judicial powers. The investigating bodies are primarily part of the *executive*, with police **analysts (users)** (S Martin) conducting the investigation, overseen by **police leadership** (S Dr. D) and also controlled by **regulators** like data protection or compliance offices. The *judiciary*, however, also plays a controlling role during **preliminary investigations** (S Robert), allowing for specific measures, for example, by issuing a warrant. Later, it manages **court proceedings** (S Muller) and questions of legality can ultimately be decided on a **constitutional** level. The third power, *legislative*, is not directly involved in investigations but sets the boundary conditions for law enforcement through regulations, usually through **parliaments**. The area of **politics**, employs a ambiguous role in this case, influencing decisions but also constituting a part of the executive. Conflicting fields can arise between all government levels. For example, the top levels might put pressure on the bottom to produce results, promoting automated analysis for its efficiency. Analysts, in turn, may use legally questionable methods, the judiciary may be concerned about the failure of legal proceedings in such cases, and regulators may be concerned about established practices that run counter to the intentions of Parliament. A recently observed problem occurs when the relationship between the system implemented as an assistant and the sovereign analyst is reversed. This can lead to effects resembling defensive decision making, where police officers intentionally make suboptimal decisions by following the results of the machine "assistants" even when they disagree. This is mostly explained by pressure from "above" and the need to protect themselves from redress if something goes wrong [9].

Software Provider — The software provider develops the tools officers use in their investigations. Here, **developers** (e.g., S Molly) implement the systems and algorithms. In doing so, it is expected that they know not only the technical details but also being aware of ethical implications. In contrast, the **management** has to mediate between the **investors/shareholders**, usually following a profit interest, the cost of implementing ethically flawless systems, and the pressure by the customers (police) to develop usable, efficient and productive systems. It is important to note that the software provider has typically no complete control over all aspects of a software system, as typically external dependencies, models, or training data are being used.

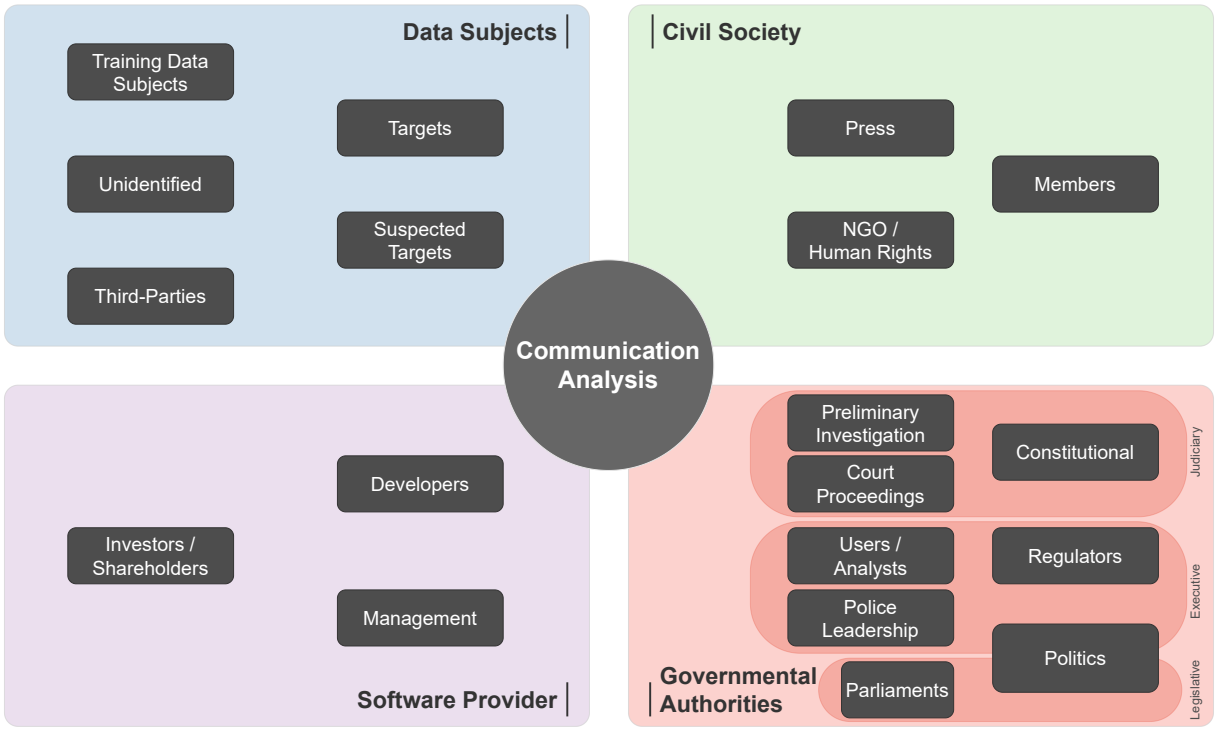


Figure 1: The stakeholders involved in communication analysis from the perspective of intelligence analysis, with conflicting interests giving rise to ethical dilemmas. We propose four main pillars of stakeholders: the *civil society*, the *governmental authorities*, the *software provider*, and the *data subjects*, each with its own subgroups of stakeholders, such as targets, developers, analysts, or NGOs.

Civil Society — Civil society can materialize not only in the form of mass media through its **members**, but also in the form of NGOs and human rights groups, arts and culture, street protests, whistleblowers, ethics councils, and so on. As such, it deliberates on what can be considered as acceptable ethical behavior in a given society, which parliament follows (through elections), and which can change over time. On the one hand, civil society can act as a corrective, for example, through critical reporting by the **press**, or legal advocacy through **NGOs or human rights groups**. Cases of unfair treatment, when entering the public agenda, can trigger the revisiting of fundamental ethical questions (as in the case of the criticism of the Northpointe recidivism algorithm and the debates it triggered about different notions of fairness and justice [43, 64, 98]). However, mass media and public deliberation can also proliferate misleading ideas about what algorithms can and cannot do. These "socio-technical imaginaries" [52] have concrete implications for how systems are being used, for the transfer and negotiation of responsibility, as well as public acceptance [13].

4 MITIGATION TECHNIQUES THROUGH VISUAL ANALYTICS

Addressing the ethical issues raised at the outset of the paper and negotiating the conflicting interest of different stakeholders is not a trivial task. Given all the different stakes involved, ethically-aware design of intelligence applications can not reasonably aim at implementing technical solutions to safeguard against all possible pitfalls. Rather it seeks to accomplish a serious consideration and balancing of the inherent trade-offs and inter-dependencies between different concerns, interests, and principles. For example, privacy-by-design may limit possibilities for advanced accountability. It thus needs to be negotiated which good is more important in each specific context and how to best achieve this. In doing so, we propose a socio-technical approach, not looking at possible technical solutions in isolation but as embedded phenomena, interacting with their environment [100]. Human interaction with technology is shaped by increasingly sophisticated and environmentally interwoven interfaces, connecting technical components, humans, and their surroundings [87]. Despite a rising appreciation of milieu-oriented approaches to understanding and designing interfaces [11], the development of interfaces has traditionally been dominated by technical disciplines and, from an ethico-political point of view, not sufficiently considered the complex socio-technical dynamics at play [40, 50]. To fill this gap, it seems most productive to work with an extended definition of the interface going beyond isolated, technocentric meanings [66, 67]. We thus adopt the proposal to focus on "interfacing" as a joint practice of "becoming-with" of humans, machines, and environments [45]. One could also call this process an intra-action [12], in which something new emerges, irreducible to its parts. Such an understanding leads to a more compehensive

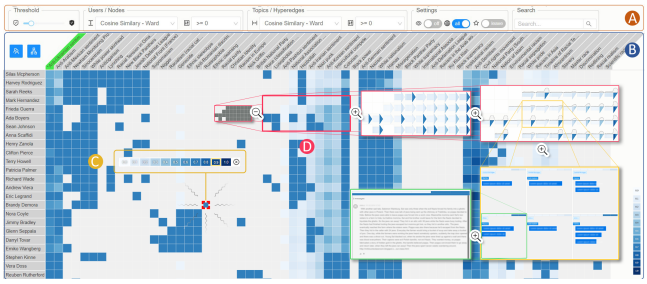


Figure 2: Example for the interactive exploration of a deep learning model. In HyperMatrix [33], a conversation topic analysis and probability framework, model predictions are visualized B through clustering and color-encoded by confidence, with interactive control elements A , semantic zooming D , and model corrections C offered for an in-depth exploration.

understanding and appreciation of what needs to be considered when designing user interfaces, especially in sensitive high-stakes areas such as communication analysis of intelligence information.

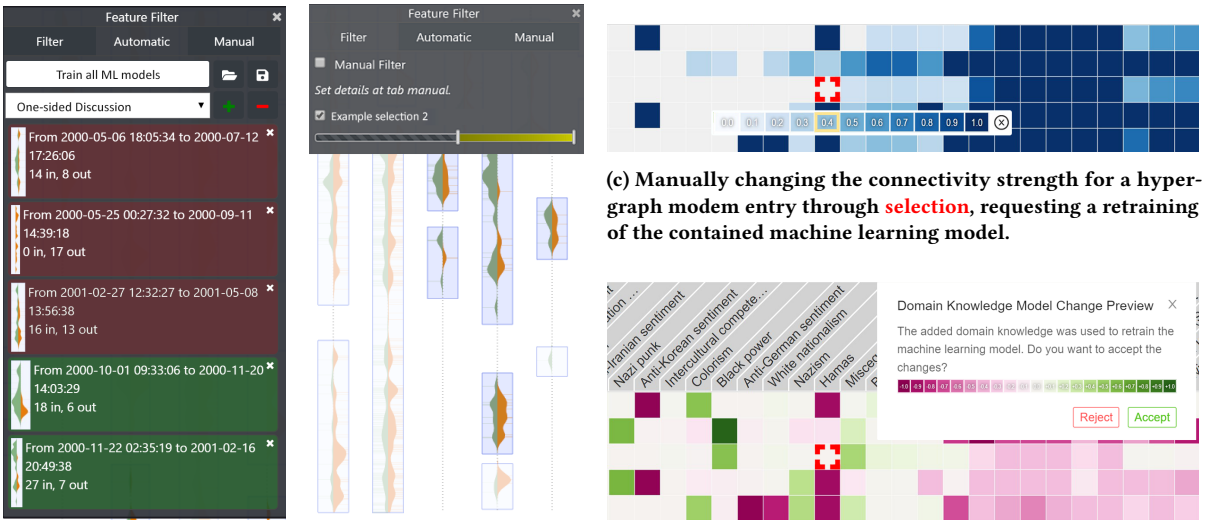
4.1 Technical Measures

In the following, we investigate common interface design methods employed in VA-based systems — the technical side – from an interfacing perspective – including the political, responsibility, intra-active, and ethical dimensions — as part of an intertwined becoming-with one another. When describing relevant aspects, we refer back to the actors from the scenario in Section 3.3. As such, we identify areas where VA, through its human agency approach, is superior to fully-automated systems while also considering the additional burdens through the distribution of responsibility.

Interactive Exploration — One key concept of VA is that instead of merely generating results, such a system supports the knowledge generation process of analysts by enabling them to learn from the data space through supported **interactions** [57]. As part of this process, ethical mitigation techniques can be integrated. In the context of communication analysis, we introduce some of the critical aspects of VA by the example of HyperMatrix [33], a conversation topic analysis and probability framework that uses a geometric deep learning approach. In our hypothetical scenario, police analyst S Martin uses the software. Instead of presenting lists of users and topics, it features an interactive design, shown in Figure 2. The developers, in our scenario S Molly , deployed a **matrix-based visualization** B for the hypergraph network structure due to increased scalability, which represents model predictions through **clustering** and **color-encoded** by confidence. The design can be considered as a form of **dimensionality reduction** [55], presenting the complex tensor model in a more comprehensible format. This supports the detection of patterns, while color-encoding facilitates pre-attentive understanding, helping S Martin to distinguish between users communicating about similar topics, like S Chris and S Carlos . Further, a multi-level visual **semantic zoom** through multiple, more detailed **in-line visualizations**, shown as insets D , allows for a more-detailed exploration, preventing an initial mental overload of S Martin . **Steering** is offered by interactive control elements A , allowing S Martin to set methods, cutoffs, and thresholds, thereby granting him agency and creativity in his usage of the system. Similarly, the system features elements from **active learning** enabling S Martin to interactively modify the model C to create something new and unique for the purpose at hand in the spirit of his analysis "becoming-with" the system. When using the system, the analyst explores the probabilities, refines model parameters, investigates hypotheses, validates change effects as part of an (indeed intra-active and) iterative analysis **feedback loop**.

The Machine Side - Analysis and Active Learning — An example of an intra-active becoming of investigator and machine is active learning. Figure 3 shows how an analyst provides labeled examples to the system improving its probabilistic accuracy. Labeling everything is tedious and time-consuming when done manually by the user. Here, **intelligent labeling** techniques can help by only requesting human input when required, relieving analysts from exploring basic or irrelevant patterns [19]. This concept can be applied **universally** to any number of **feedback mechanisms** between system and user, affecting data selection, machine learning models, heuristic algorithms, or their parameters. Through active learning, S Martin can integrate its experience-saturated as well as its domain-specific expert knowledge into the analysis process, thereby sharpening the analysis result with regard to the field's unique requirements. Methods such as color coding imply a form of nudging through preattentive processing. Transparency, in turn, can be conveyed by visualizing consequences and effects of the active learning approach (e.g., which entries are subsequently classified differently and by how much). This corresponds to a "what-if" preview , which supports the selection process, but can also act as a "control", e.g., against unintended side effects.

The Human Side - Guidance and Explainability — Guidance describes the interplay between system and user actions and their understanding in the context of machine learning, explainable artificial intelligence (XAI), and knowledge generation. One form of guidance can manifest by the system to nudge the user in the right direction, for example, by showing similar matches or conflicting options. In the context of learning and teaching, it exhibits a wider dynamic, encompassing **system teaching**, **user teaching**, **system learning**, and **user learning**. As a process, it can be described by the knowledge generation model [57, 89] and by the co-adaptive guidance process [96, 97]. Different forms of guidance can be achieved by a **visually abstract** visualization as well as **conceptual user interaction** design. For visualization, **abstract representations** like glyph [39] can be used for improved recognizability and comparability. In communication analysis, commonly used representations are text, highlighting, concept extraction, and network display. However, depending on the individual needs, they may not suffice. During the design, the aim of the **representation**, the selection of the appropriate **visualization technique**, the **visual variables** employed, and the **color-schemes** used has to be considered. **Inherent biases** can play an essential role, affecting values as social biases (e.g., homogeneity bias), actions (e.g.,



(a) Manually supplying positive (green) and negative (red) examples to the system during active learning.

(b) Classification model applied to communication episodes after active learning improved filtering.

(c) Manually changing the connectivity strength for a hypergraph modem entry through selection, requesting a retraining of the contained machine learning model.

(d) Visualized changes in the model prediction (from negative to no to positive change). Before the changes are either rejected or accepted, the system nudges the analyst to examine the results for unintended side effects.

Figure 3: Examples of active learning in Conversational Dynamics [93] (a, b) and HyperMatrix [33] (c, d).

blind spot or Ostrich effect), and **perceptions** (e.g., illusions or Weber-Fachner-Law) [25], both during development as well as usage. For interaction, instead of filtering communication texts by keywords (the selection of which might be biased or incomplete), a **visual query language** [35] can be used where conditions are based on concepts. Here, the system may similarly suggest additional concepts for differentiation or indicate a too restricted search. In implementing such interfaces, care has to be taken regarding a neutral representation while considering the levels of detail and abstraction [53]. Too much abstraction can lead to a loss of context.

Provenance — As systems become more complex and the number of available interactions increases, the number and sequence of necessary steps during the investigation expands rapidly, making it increasingly difficult to fathom and explain them after the fact. In this context, while many systems focus on supporting the process of knowledge generation, few also emphasize how this process is carried out. [62, 68]. For reproducibility, which is crucial for accountability (see Section 4.2), knowing how analysts like S Martin **used** the system to draw conclusions is essential, but this becomes increasingly complicated when iterative and intra-active processes are involved. Instead of a **linear timeline** of steps or a **list** of explored hypotheses, VA applications can store interaction chronicles **in context** with the data, remembering **settings**, **views**, and **actions**. Through techniques like time-stamping, hashing, and digitally signing information, the chain of evidence becomes proofable. As such, they can provide tamper-proof **tracking** as well as **replay** functionality to revisit intermediate steps, while also enabling approaches like **provenance graphs** as shown in Figure 4, strengthening the chain of evidence for both analysts and subjects.

4.2 Balancing Advantages and Risks Through the Interface

In the following we identify advantages but also the limitations and risks involved in using VA for analysis of intelligence data. To do so, we refer back to the various challenges and conflicts identified in Sections 2 and 3.

A1. User Agency — First and foremost, VA approaches support user agency and help prevent blind obedience to machine results, which is certainly one of the most common dangers in the face of exaggerated expectations. VA levels the playing field between system and investigator by creating a con-genial joint agency that facilitates a sense of growing-with each other rather than replacing each other. This is due to an approach - inherent in VA systems - that is collaborative rather than hierarchical. Furthermore, rather than encouraging an either-or decision process in cases where human analysts tend to deviate from the system's proposed outcomes, potentially leading to defensive or otherwise suboptimal decision making, **explainability** can help analysts like S Martin question their own inputs that influence the outcome during active learning. This can ultimately help identify dead ends in the line of inquiry.

A2. Privacy — Guidance allows circumventing privacy issues related to uninvolved third parties because selective presentation allows outsourcing to the system, such that the private life of Martin's girlfriend S Sarah, for example, does not need to be reviewed by natural persons.

A3. Fairness — Using VA as a means to handle heterogeneous data has advantages as opposed to automated systems exclusively trained on past criminal records. The latter frequently transport racial, gender, and other harmful stereotypes into the present and

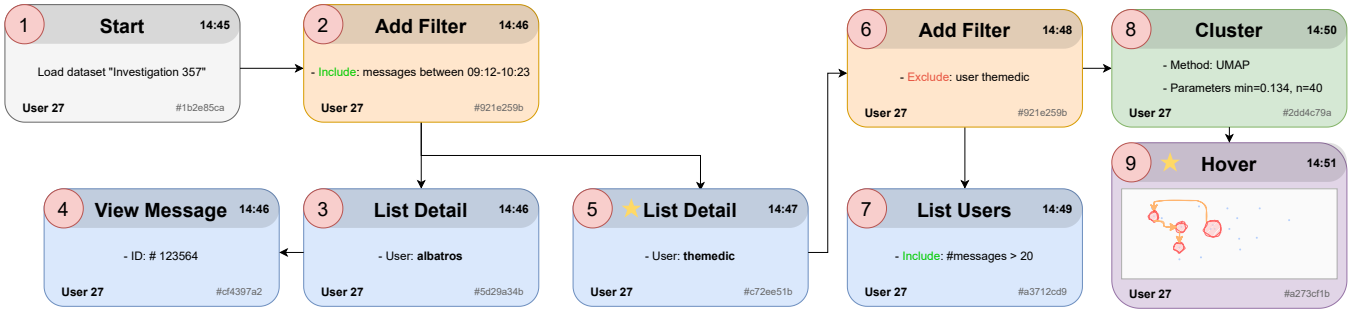


Figure 4: Example sketch of a provenance history component using a directed acyclic graph (DAG) approach with inline details instead of a linear history, allowing for a more complete picture of the explored steps (1-9) and reverts from dead-ends (3,4,7).

future, thus cementing historically grown structures of domination [69]. With VA, the analyst can inspect the reasoning and intervene with his domain knowledge through corrective actions.

A4. Efficiency — Big Data-based VA can help to visualize complex operating dynamics, thus helping to see previously undetectable correlations [73]. VA systems may provide helpful **abstractions** and powerful **analytical capabilities** to efficiently find the non-trivial needle in the haystack, supporting investigators identifying even faint tracks. However, each design tailored for more efficiency has to be evaluated for oversimplification or risks of misinterpretation.

A5. Literacy — VA encourages users to engage with and appropriate the systems they use actively and creatively, developing advanced literacy in their usage through daily practice. This literacy, then again, can be shared among colleagues, for example, by the integrated **sharing** of **recipes** through VA systems (e.g., *Common actions here are…*).

A6. Customization — Active learning supports the creation of tailored solutions instead of one-size-fits-all ones. The latter promise rather poor performance in the fight against organized crime, which requires highly localized and specialized approaches as well as detailed **domain knowledge** of experienced and highly qualified experts [83]. This is not a coincidence, but part of organized crime's recipe for success, which deliberately operates according to a nonlinear, swarm-like logic to confuse investigators and prevent a pattern-based approach from being applied. In such cases, the involvement of highly experienced experts is essential. Here, technical solutions that **combine** efficiency-enhancing **automation** with **human** criminological **expertise** to produce tailored solutions promise real gains. Furthermore, it is crucial that systems are capable of adapting dynamically to changing circumstances and environments.

4.3 Risks, Limitations and Additional Measures Required

R1. Lack of Accountability — Reproducibility is crucial for accountability (C6). This is one of the biggest challenges when considering that in instances of interfacing such as in VA, it is nearly impossible to clearly define and delineate who is responsible in cases of potentially flawed output and decisions since these need to be considered as the result of a joint becoming-with one another of the analysts' knowledge and the system's analysis. Apart from the exact *software versions*, the *starting seeds* of pseudo-random generators for non-deterministic algorithms, the exact *data used in the training process*, what is needed is also detailed information about how analysts like S Martin **interacted** with the system. Just presenting the final analysis is certainly not enough in such cases, as the decision-making process of S Martin was not strictly linear. Learning has occurred along the way, and a judge like S Muller wants to be satisfied that all viable hypotheses have been explored, and were not merely neglected. Advanced **provenance** approaches in VA designs are needed to provide this accountability.

R2. Training and Community-Building Among Users — It can also not be taken for granted that police officers possess the necessary background knowledge to understand the inner workings of the applied ML and VA systems, their potential shortcomings, and continual development. On the one hand, developers like S Molly have to mobilize considerable efforts to design explanations and train users in the effective and critical use of the systems. The way forward will be to design such training not as a one-way transfer of how-to-use information, but to curate it as an interactive process between software providers and analysts like S Martin, as well as *between* the analysts themselves (e.g., through design studies). Possible directions are the privacy-preserving use of **gamification** [90] inside **VA system**, where users can compete in practical challenges, or online discussion **forums**, where users can highlight and socially negotiate limitations and pitfalls. Moreover, the VA system and training activities have to be sensible to the shift of power structures triggered by the introduction of new software [47]. Young police officers with considerable digital literary will likely be favored by the transformation of work practices, whereas long-standing investigators might feel marginalized [56]. To avoid losing valuable experiential knowledge of senior investigators, it seems reasonable to work with technical solutions that make it possible to include rather than exclude these qualitative dimensions.

R3. Prevent Automated Inequality — Individual and institutional racism can be infiltrated into the system through active learning, with individuals influencing the system's learning process with their often unconscious biases [25]. This is a significant risk in the context of collaborative solutions such as VA, which grant investigators and officers increased responsibility in the training and design of automated analysis systems. This risk must be kept

in mind and urgently requires further control mechanisms, especially in light of the alarming statistics and research on racial and sexist prejudices (unconsciously) harbored and enacted by many police officers [32, 37, 59, 86]. Active learning and other **human in command** solutions that result in automation based on user input, like from S Martin, should therefore only be employed if they have been preceded by in-depth anti-discrimination training and education. The same argument may be applied to developers like S Molly, whose biases and inherently (wrong) assumptions may be introduced by an inadequate design of the system. Considering and integrating adequate mitigation measures already at the VA system level may prove especially fruitful, as the invested time is regained through the disseminator effect.

R4. Facilitating critical reflection — The system and user interface should incentivise instances of reflection [15]. On the one hand, these instances of reflection (e.g., through text-based nudges and explanations) can help users recognize their own unconscious biases while interacting with the system, e.g., by showing warning signs after allegedly racist or sexist search queries or by blocking such queries all along (supplemented by an explanation). [102] On the other hand, such reflective features should enable users to challenge outputs and decisions of the system and support a critical use of the software by the investigators. **Reflective design** [95] can, accordingly, both inform the active learning process and improve the technical literacy of police investigators to interpret and evaluate system results.

R5. Human Oversight — Because of the risks mentioned above, approaches that support shared responsibility and foster a joint agency between analysts and systems must be accompanied by additional instances of human oversight. At a minimum, there needs to be a regular **review** of whether systems are being trained during the course of use and continuous input to ensure fair treatment, especially of protected groups. In addition, technical mitigation strategies can be considered. A VA **provenance** system that effectively negotiates responsibilities can remember the individual agents who interacted with the system or made changes through active learning. From a privacy perspective, it needs to secure and protect the identity of these agents but also allow for discussion of problematic decisions among peers and authorities (i.e., specific police officers who have labeled a particular entity differently). This would enable leadership like S Dr. D to validate in a privacy-preserving way how much biases occur within their organization. Caution, however, must be taken not to arrive at a culture of control, where long-serving officers feel deprived of their agency [56]

4.4 Negotiating Risks and Advantages Through the Interface

VA systems present a particular instance of interface-oriented solutions. As such they conform to data infrastructure platforms, allowing acquiring, handling, leveraging, and storing information and accompanying metadata from heterogeneous sources. Hidden biases in fully automated systems trained on biased past crime data or otherwise problematic, dirty, or insufficient big data sets can be mitigated through more dynamic, real-time, and interactive approaches. However, discriminatory biases can also be reinforced through VA-methods, placing much responsibility on users. Therefore, in-process measures of fairness are urgently needed. To this end, it is absolutely necessary to always ensure the traceability of the process genesis, enable accountability in cases of abuse, and to initiate retraining processes accordingly. The transparency required must also be harmonized with privacy requirements, which presupposes a certain level of abstraction and can only work well with decently trained personnel. In contrast to the manual combination of separate analyses or fully automated approaches, VA can be designed with built-in consideration of ethical issues for the entire knowledge generation process, adjusting user expectations to technical capabilities (C4), counteracting opacities (C3) and meeting privacy requirements (C2). Through human oversight (C5) by different stakeholders, discrimination (C1) can be detected. Simultaneously, by automatically collecting tamper-proof provenance about the system and the human, VA systems can increase trustworthiness and accountability (C6).

5 CONCLUSION

We have shown in detail how various VA methods can address ethical challenges in advanced analytical systems. While we have highlighted concrete points to consider, we do not intend to present a fixed set of rules for the ethically conscious design of VA systems. Indeed, in our view, this is always a matter of negotiating trade-offs between conflicting interests, which can vary widely depending on the unfolding interaction dynamics. Therefore, we aim to stimulate a discussion about the consideration of ethical implications as an integral part of the design process from the outset.

Although we focus on the case of intelligence applications, many of the results of our work and the ethical discussion are more generally applicable to the design of VA applications. This is because, on a more abstract level, our approach leads us to the insight that one of the main advantages of VA methods is that they take the *interface* as a starting point for technological innovation. This means that innovation is approached not only in a technical, but rather in a **socio-technical** way. Useful innovations cannot exist in isolation and should pay attention to their impact on communities and society as a whole. This shifts the focus to embedding technologies into existing social institutions such as police departments and criminal justice agencies. Here, new technologies are most useful when combining the benefits of efficiency-enhancing automation with the experience-saturated knowledge of investigators. Concentrating on the interface has obvious advantages, as it implies a focus on the situated nature of human-computer-configurations.

By capitalizing on instead of trying to stabilize the potential openness, and thus eventfulness of technological interconnectedness, undesirable dynamics can be dealt with much more proactively, while flexibly addressing the situational and individual needs of different cases. The issue is not only what interacts with whom, but also how new phenomena emerge as part of complex intra-active configurations of people and automation systems. Discriminating behavior of individual investigators, for example, might not affect the functioning of the police force as a whole. When multiplied and cemented in automation processes, however, it can contribute to structural discrimination and patterns of unfair treatment on a larger scale. By responding transparently to investigator input, VA

can help well-meaning analysts recognize their often unconscious biases by showing them how harmful social stereotypes sometimes cause them to overlook features and stumble down blind alleys.

To avoid risks and increase benefits, it is important to keep in mind that interfacing is not simply the matching of two separate entities, but the creation of something fundamentally new, a hybrid of human and machine. This hybrid requires tailored quality insurance measures such as adapted training, new forms of oversight involving technical, legal, and ethical experts, and also adapted policy and ethical frameworks that focus not solely on the technologies or on the user, but on what emerges as something new in the interaction between these entities.

ACKNOWLEDGMENTS

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany (BMBF) in the framework of PEGASUS under the program "Forschung für die zivile Sicherheit 2018 - 2023" and its announcement "Zivile Sicherheit - Schutz vor organisierter Kriminalität II". The BMBF had no role in the design and conduct of the analysis or the preparation, review, or approval of the manuscript. The authors declare no other financial interests.

REFERENCES

[1] AI Ethics Impact Group. 2020. From Principles to Practice An Interdisciplinary Framework to Operationalise AI Ethics. https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WKIO_2020_final.pdf

[2] Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. 2016. Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective. *PeerJ Computer Science* 2 (2016), e93. https://doi.org/10.7717/peerj-cs.93

[3] American Civil Liberties Union. 2016. Statement of Concern About Predictive Policing by ACLU and 16 Civil Rights Privacy, Racial Justice, and Technology Organizations. https://www.aclu.org/other/statement-concern-about-predictive-policing-aclu-and-16-civil-rights-privacy-racial-justice

[4] Morgan G. Ames. 2015. Charismatic Technology. *Aarhus Series on Human Centered Computing* 1, 1 (2015), 12. https://doi.org/10.7146/aahcc.v1i1.21199

[5] Morgan G. Ames. 2019. *The Charisma Machine: The Life, Death, and Legacy of One Laptop per Child.* MIT Press, Cambridge, MA, USA.

[6] Louise Amoore. 2014. Security and the Claim to Privacy. *International political sociology* 8, 1 (2014), 108–112. http://dro.dur.ac.uk/14920/

[7] Louise Amoore. 2019. Doubt and the Algorithm: On the Partial Accounts of Machine Learning. *Theory, Culture & Society* 36, 6 (2019), 147–169. https://doi.org/10.1177/0263276419851846

[8] Mike Ananny and Kate Crawford. 2018. Seeing Without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability. *New Media & Society* 20, 3 (2018), 973–989. https://doi.org/10.1177/1461444816676645

[9] Florian M. Artinger, Sabrina Artinger, and Gerd Gigerenzer. 2019. C. Y. A.: Frequency and Causes of Defensive Decisions in Public Administration. *Business Research* 12, 1 (2019), 9–25. https://doi.org/10.1007/s40685-018-0074-2

[10] Jef Ausloos. 2012. The 'Right to be Forgotten' – Worth remembering? *Computer Law & Security Review* 28, 2 (2012), 143–152. https://doi.org/10.1016/j.clsr.2012.01.006

[11] Ciano Aydin, Margoth González Woge, and Peter-Paul Verbeek. 2019. Technological Environmentality: Conceptualizing Technology as a Mediating Milieu. *Philosophy & Technology* 32, 2 (2019), 321–338. https://doi.org/10.1007/s13347-018-0309-3

[12] Karen Barad. 2014. Diffracting Diffraction: Cutting Together-Apart. *Parallax* 20, 3 (2014), 168–187. https://doi.org/10.1080/13534645.2014.927623

[13] Jascha Bareis and Christian Katzenbach. 2021. Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics. *Science, Technology & Human Values* (2021), 01622439211030007. https://doi.org/10.1177/01622439211030007

[14] Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2017. Fairness and Machine Learning. *Nips Tutorial* 1 (2017).

[15] Eric P.S. Baumer. 2015. Reflective Informatics: Conceptual Dimensions for Designing Technologies of Reflection. In *Proceedings of the 33rd CHI Conference on Human Factors in Computing Systems (CHI)*, Bo Begole, Jinwoo Kim, Kori Inkpen, and Woontack Woo (Eds.). ACM, New York, NY, USA, 585–594. https://doi.org/10.1145/2702123.2702234

[16] Ruha Benjamin. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code.* Polity, Medford MA.

[17] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. 1999. *Readings in Information Visualization: Using Vision to Think.* Morgan Kaufmann, San Francisco, CA.

[18] John M. Carrol. 1999. Five Reasons for Scenario-Based Design. In *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences*, Ralph H. Sprague (Ed.). IEEE Comput. Soc, 11. https://doi.org/10.1109/HICSS.1999.772890

[19] Michael Correll. 2019. Ethical Dimensions of Visualization Research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Stephen Brewster, Geraldine Fitzpatrick, Anna Cox, and Vassilis Kostakos (Eds.). ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300418

[20] Sasha Costanza-Chock. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need.* MIT Press, New York. https://library.oapen.org/handle/20.500.12657/43542

[21] DataWalk Inc. 2020. DataWalk. https://datawalk.com/

[22] Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. Unique in the Crowd: The Privacy Bounds of Human Mobility. *Scientific Reports* 3 (2013), 1376.

[23] Erinç Dikici, Maël Fabien, Jan Hornek, Joshua Hughes, Miroslav Janošik, Marek Kovac, Petr Motlicek, Hoang H. Nguyen, Shantipriya Parida, Johan Rohdin, et al. 2021. ROXSD: A Simulated Dataset of Communication in Organized Crime. *ISCA Symposium on Security and Privacy in Speech Communication* (2021), 32–36.

[24] Mennatallah El-Assady, Valentin Gold, Markus John, Thomas Ertl, and Daniel A. Keim. 2016. Visual Text Analytics in Context of Digital Humanities. In *1st IEEE VIS Workshop on Visualization for the Digital Humanities (Vis4DH) as part of the IEEE VIS 2016.* https://scibib.dbvis.de/publications/view/686

[25] Geoffrey Ellis (Ed.). 2018. *Cognitive Biases in Visualizations* (1st ed. 2018 ed.). Springer International Publishing and Imprint: Springer, Cham.

[26] Virginia Eubanks. 2018. *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor* (illustrated edition ed.). St Martin's Press.

[27] European Commission. 2021. White Paper on Artificial Intelligence - A European Approach to Excellence and Trust. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0065&from=EN

[28] European Commission. 2021-04-21. Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM/2021/206 final): Artifical Intelligence Act. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

[29] European Commission. 2021-12-09. Proposal for a Directive of the European Parliament and of the Council on Improving Working Conditions in Platform Work (COM(2021) 762). https://ec.europa.eu/social/BlobServlet?docId=24992&langId=en

[30] European Parliament and the European Council. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation): GDPR.

[31] Europol. 2021. European Union Serious and Organised Crime Threat Assessment (SOCTA): A Corrupting Influence: The Infiltration and Undermining of Europe's Economy and Society by Organised Crime. https://www.europol.europa.eu/cms/sites/default/files/documents/socta2021_1.pdf

[32] Jeffrey Fagan, Anthony A. Braga, Rod K. Brunson, and April Pattavina. 2016. Stops and Stares: Street Stops, Surveillance, and Race in the New Policing. *Fordham Urban Law Journal* 43, 3 (2016), 77.

[33] Maximilian T. Fischer, Devanshu Arya, Dirk Streeb, Daniel Seebacher, Daniel A. Keim, and Marcel Worring. 2020. Visual Analytics for Temporal Hypergraph Model Exploration. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 550–560. https://doi.org/10.1109/TVCG.2020.3030408

[34] Maximilian T. Fischer, Frederik Dennig, Daniel Seebacher, Daniel A. Keim, and Mennatallah El-Assady. 2021. Communication Analysis through Visual Analytics: Current Practices, Challenges, and New Frontiers. arxiv:2106.14802

[35] Maximilian T. Fischer, Daniel Seebacher, Rita Sevastjanova, Daniel A. Keim, and Mennatallah El-Assady. 2021. CommAID: Visual Analytics for Communication Analysis through Interactive Dynamics Modeling. *Computer Graphics Forum* 40, 3 (2021), 25–36. https://doi.org/10.1111/cgf.14286

[36] Sofie Flensburg and Signe Sophus Lai. 2020. Mapping Digital Communication Systems: Infrastructures, Markets, and Policies as Regulatory Forces. *Media, Culture & Society* 42, 5 (2020), 692–710. https://doi.org/10.1177/0163443719876533

[37] Cortney A. Franklin. 2005. Male Peer Support and the Police Culture. *Women & Criminal Justice* 16, 3 (2005), 1–25. https://doi.org/10.1300/J012v16n03{_}01

[38] William J. Frawley, Gregory Piatetsky-Shapiro, and Christopher J. Matheus. 1992. Knowledge Discovery in Databases: An Overview. *AI Magazine* 13, 3 (1992), 57. https://doi.org/10.1609/aimag.v13i3.1011

[39] Johannes Fuchs, Petra Isenberg, Anastasia Bezerianos, and Daniel A. Keim. 2017. A Systematic Review of Experimental Studies on Data Glyphs. *IEEE Transactions on Visualization and Computer Graphics* 23, 7 (2017), 1863–1879. https://doi.org/10.1109/TVCG.2016.2549018

[40] Alexander R. Galloway. 2012. *The Interface Effect.* Wiley.

[41] Clare Garvie and Jonathan Frankle. 2016. Facial-Recognition Software Might Have a Racial Bias Problem. *The Atlantic* (2016). https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/

[42] Lisa Gitelman. 2013. *Raw Data is an Oxymoron.* The MIT Press, Cambridge, Massachusetts.

[43] Ben Green. 2018. Putting the J(ustice) in FAT. https://medium.com/berkman-klein-center/putting-the-j-ustice-in-fat-28da2b8eae6d

[44] Rachid Hadjidj, Mourad Debbabi, Hakim Lounis, Farkhund Iqbal, Adam Szporer, and Djamel Benredjem. 2009. Towards an Integrated E-mail Forensic Analysis Framework. *Digital Investigation* 5, 3-4 (2009), 124–137. https://doi.org/10.1016/j.diin.2009.01.004

[45] Donna J. Haraway. 2017. *Staying with the Trouble: Making Kin in the Chthulucene. Experimental Futures* (illustrated edition ed.). Academic Press.

[46] Paula Helm and Thilo Hagendorff. 2021. Beyond the Prediction Paradigm: Challenges for AI in the Struggle Against Organized Crime. *Law and Contemporary Problems* 84, 3 (2021), 1–17.

[47] Kathryn Henderson. 1991. Flexible Sketches and Inflexible Data Bases: Visual Communication, Conscription Devices, and Boundary Objects in Design Engineering. *Science, Technology & Human Values* 16, 4 (1991), 448–473. http://sth.sagepub.com/content/16/4/448.short

[48] Matthew Honnibal. 2019. SpaCy: An NLP library. https://www.spacy.io

[49] Chris Jay Hoofnagle, Bart van der Sloot, and Frederik Zuiderveen Borgesius. 2018. The European Union General Data Protection Regulation: What it is and What it Means. *Information & Communications Technology Law* 28, 1 (2018). https://doi.org/10.1080/13600834.2019.1573501

[50] Branden Hookway. 2014. *Interface.* MIT Press.

[51] IBM. 2020. i2 Analyst's Notebook. https://www.ibm.com/us-en/marketplace/analysts-notebook

[52] Sheila Jasanoff. 2015. Future Imperfect: Science, Technology, and the Imaginations of Modernity. In *Dreamscapes of Modernity.* University of Chicago Press, 1–33. https://doi.org/10.7208/9780226276663-001

[53] Wolfgang Jentner, Rita Sevastjanova, Florian Stoffel, Daniel A. Keim, Jürgen Bernard, and Mennatallah El-Assady. 2018. Minions, Sheep, and Fruits : Metaphorical Narratives to Explain Artificial Intelligence and Build Trust. In *Workshop on Visualization for AI Explainability at IEEE.* https://bib.dbvis.de/uploadedFiles/minionssheepfruits.pdf

[54] Elizabeth E. Joh. 2017. Feeding the Machine: Policing Crime Data & Algorithms. *William & Mary Bill of Rights Journal* 26 (2017), 287.

[55] Samuel Kaski and Jaakko Peltonen. 2011. Dimensionality Reduction for Data Visualization. *IEEE Signal Processing Magazine* 28, 2 (2011), 100–104. https://doi.org/10.1109/MSP.2010.940003

[56] Mareile Kaufmann, Simon Egbert, and Matthias Leese. 2018. Predictive Policing and the Politics of Patterns. *The British Journal of Criminology* 59, 3 (2018), 674–692. https://doi.org/10.1093/bjc/azy060

[57] Daniel A. Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. 2008. Visual Analytics: Definition, Process, and Challenges. In *Information Visualization*, Andreas Kerren (Ed.). Lecture Notes in Computer Science, Vol. 4950. Springer, 154–175. https://doi.org/10.1007/978-3-540-70956-5-7

[58] Daniel A. Keim, Jörn Kohlhammer, Geoffrey P. Ellis, and Florian Mansmann. 2010. *Mastering the Information Age Solving Problems with Visual Analytics.* Eurographics Association.

[59] Stefanie Kemme, Iniobong Essien, and Marleen Stelter. 2020. Antimuslimische Einstellungen in der Polizei?: Der Zusammenhang von Kontakthäufigkeit und -qualität mit Vorurteilen und Stereotypen gegenüber Muslimen. *Monatsschrift für Kriminologie und Strafrechtsreform* 103, 2 (2020), 129–149. https://doi.org/10.1515/mks-2020-2048

[60] Robert O. Keohane and Joseph S. Nye Jr. 1998. Power and Interdependence in the Information Age. *Foreign Affairs* 77 (1998), 81.

[61] Rob Kitchin. 2017. Thinking Critically About and Researching Algorithms. *Information, Communication & Society* 20, 1 (2017). https://doi.org/10.1080/1369118X.2016.1154087

[62] Robert Kosara and Jock Mackinlay. 2013. Storytelling: The Next Step for Visualization. *Computer* 46, 5 (2013), 44–50. https://doi.org/10.1109/MC.2013.36

[63] Keita Kurita, Paul Michel, and Graham Neubig. 2020. Weight Poisoning Attacks on Pretrained Models. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Stroudsburg, PA, USA, 2793–2806. https://doi.org/10.18653/v1/2020.acl-main.249

[64] Jeff Larson and Julia Angwin. 2016. Technical Response to Northpointe. *ProPublica* (2016). https://www.propublica.org/article/technical-response-to-northpointe

[65] Bruno Lepri, Nuria Oliver, Emmanuel Letouzé, Alex Pentland, and Patrick Vinck. 2018. Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges. *Philosophy & Technology* 31, 4 (2018), 611–627. https://doi.org/10.1007/s13347-017-0279-x

[66] Benjamin Lipp. 2017. Analytik des Interfacing. Zur Materialität technologischer Verschaltung in prototypischen Milieus robotisierter Pflege. *1866-2447* 10 (2017), 107–129. https://doi.org/10.6094/behemoth.2017.10.1.948

[67] Benjamin Lipp, Paula Helm, Athanasios Karafillidis, and Roser Pujadas. 2022. Theorising Interfaces / Interface Theories. In *EASST 2022.*

[68] Yang Liu, Tim Althoff, and Jeffrey Heer. 2020. Paths Explored, Paths Omitted, Paths Obscured: Decision Points & Selective Reporting in End-to-End Data Analysis. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Regina Bernhaupt, Florian 'Floyd' Mueller, David Verweij, Josh Andres, Joanna McGrenere, Andy Cockburn, Ignacio Avellino, Alix Goguey, Pernille Bjørn, Shengdong Zhao, Briane Paul Samson, and Rafal Kocielnik (Eds.). ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376533

[69] Kristian Lum and William Isaac. 2016. To Predict and Serve? *Significance* 13, 5 (2016), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

[70] Adrian Mackenzie. 2015. The Production of Prediction: What Does Machine Learning Want? *European Journal of Cultural Studies* 18, 4/5 (2015), 429–445.

[71] Kevin Macnish. 2018. *The Ethics of Surveillance: An Introduction.* Routledge Taylor & Francis Group, London and New York.

[72] Christopher D. Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing.* MIT Press, Cambridge, Mass. and London.

[73] Viktor Mayer-Schönberger and Kenneth Cukier. 2013. *Big Data: Die Revolution, die unser Leben verändern wird* (1. ed ed.). Redline-Verlag, München.

[74] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys* 54, 6 (2021), 1–35. https://doi.org/10.1145/3457607

[75] Jacob Metcalf and Kate Crawford. 2016. Where are Human Subjects in Big Data Research? The Emerging Ethics Divide. *Big Data & Society* 3, 1 (2016), 2053951716650211. https://doi.org/10.1177/2053951716650211

[76] Christoph Molnar. 2019. *Interpretable Machine Learning: A Guide for Making Black Box Models Interpretable.* Lulu, Morisville, North Carolina.

[77] Anouk Mols and Susanne Janssen. 2017. Not Interesting Enough to be Followed by the NSA. *Digital Journalism* 5, 3 (2017), 277–298. https://doi.org/10.1080/21670811.2016.1234938

[78] Helen Nissenbaum. 1994. Computing and accountability. *Communications of the ACM* 37, 1 (1994), 72–80. https://doi.org/10.1145/175222.175228

[79] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York University Press, New York.

[80] Nuix Pty Ltd. 2020. Nuix Discover and Nuix Investigate. https://www.nuix.com/products

[81] Cathy O'Neil. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (1st edition ed.). Penguin Books UK.

[82] Palantir Technologies, Inc. 2020. Gotham. https://www.palantir.com/palantir-gotham/

[83] Letizia Paoli. 2002. The Paradoxes of Organized Crime. *Crime, Law and Social Change* 37, 1 (2002), 51–97. https://doi.org/10.1023/A:1013355122531

[84] Frank Pasquale. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information.* Harvard University Press.

[85] Judy C. Pearson. 2011. *Human Communication* (4th ed.). McGraw-Hill, New York.

[86] E. Ashby Plant and B. Michelle Peruche. 2005. The Consequences of Race for Police Officers' Responses to Criminal Suspects. *Psychological Science* 16, 3 (2005), 180–183. https://doi.org/10.1111/j.0956-7976.2005.00800.x

[87] Roser Pujadas, Erika Valderrama, and Will Venters. 2020. Interfaces and the Dynamics of Digital Ecosystems: A Study of the Online Travel Ecosystem. *International Conference on Information Systems (ICIS)* (2020).

[88] Cynthia Rudin. 2019. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence* 1, 5 (2019), 206–215. https://doi.org/10.1038/s42256-019-0048-x

[89] Dominik Sacha, Andreas Stoffel, Florian Stoffel, Bum Chul Kwon, Geoffrey P. Ellis, and Daniel A. Keim. 2014. Knowledge Generation Model for Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1604–1613. https://doi.org/10.1109/TVCG.2014.2346481

[90] Michael Sailer, Jan Hense, J Mandl, and Markus Klevers. 2014. Psychological perspectives on motivation through gamification. *Interaction Design and Architecture Journal* 19 (2014), 28–37.

[91] Carlos Alberto Scolari. 2009. Mapping Conversations About New Media: The Theoretical Field of Digital Communication. *New Media & Society* 11, 6 (2009), 943–964. https://doi.org/10.1177/1461444809336513

[92] John Scott. 2017. *Social Network Analysis* (4th ed.). SAGE, Los Angeles.

[93] Daniel Seebacher, Maximilian T. Fischer, Rita Sevastjanova, Daniel A. Keim, and Mennatallah El-Assady. 2019. Visual Analytics of Conversational Dynamics. In *EuroVis Workshop on Visual Analytics (EuroVA) (EuroVA)*, Tatiana von Landesberger and Cagatay Turkay (Eds.). The Eurographics Association, Porto,

Portugal. https://doi.org/10.2312/eurova.20191130

[94] Andrew D. Selbst. 2018. Disparate Impact in Big Data Policing. *Georgia Law Review* 52, 1 (2018), 3373.

[95] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph 'Jofish' Kaye. 2005. Reflective Design. In *Proceedings of the 4th Decennial Conference on Critical Computing Between Sense and Sensibility - CC '05*. ACM Press, Aarhus, Denmark, 49. https://doi.org/10.1145/1094562.1094569

[96] Fabian Sperrle, Astrik Jeitler, Jürgen Bernard, Daniel Keim, and Mennatallah El-Assady. 2021. Co-Adaptive Visual Data Analysis and Guidance Processes. *Computers & Graphics* 100 (2021), 93–105. https://doi.org/10.1016/j.cag.2021.06.016

[97] Fabian Sperrle, Astrik Jeitler, Jürgen Bernard, Daniel A. Keim, and Mennatallah El-Assady. 2020. Learning and Teaching in Co-Adaptive Guidance for Mixed-Initiative Visual Analytics. In *EuroVis Workshop on Visual Analytics (EuroVA)*, Cagatay Turkay and Katerina Vrotsou (Eds.). https://doi.org/10.2312/eurova.20201088

[98] Matthias Spielkamp. 2017. Inspecting Algorithms for Bias. *MIT Technology Review* (2017). https://www.technologyreview.com/2017/06/12/105804/inspecting-algorithms-for-bias/

[99] Florian Stoffel, Wolfgang Jentner, Michael Behrisch, Johannes Fuch, and Daniel A. Keim. 2017. Interactive Ambiguity Resolution of Named Entities in Fictional Literature. *Computer Graphics Forum* 36, 3 (2017), 189–200.

[100] Lucille Alice Suchman. 2007. *Human-Machine Reconfigurations: Plans and Situated Actions* (2nd ed.). Cambridge University Press, Cambridge.

[101] Wouter van Atteveldt and Tai-Quan Peng. 2018. When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science. *Communication Methods and Measures* 12, 2-3 (2018), 81–92. https://doi.org/10.1080/19312458.2018.1458084

[102] Emily Wall, John Stasko, and Alex Endert. 2019. Toward a Design Space for Mitigating Cognitive Bias in Vis. In *2019 IEEE Visualization Conference (VIS)*. IEEE, 111–115. https://doi.org/10.1109/VISUAL.2019.8933611

[103] Gregor Wiedemann, Seid Muhie Yimam, and Chris Biemann. 2018. New/s/leak 2.0 – Multilingual Information Extraction and Visualization for Investigative Journalism. In *Social Informatics*, Steffen Staab, Olessia Koltsova, and Dmitry I. Ignatov (Eds.). Lecture Notes in Computer Science, Vol. 11186. Springer International Publishing, Cham, 313–322. https://doi.org/10.1007/978-3-030-01159-8{_}30

[104] Tal Zarsky. 2016. The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science, Technology & Human Values* 41, 1 (2016), 118–132. https://doi.org/10.1177/0162243915605575