EuroVis Workshop on Visual Analytics (2026)
D. Archambault and A. Vilanova (Editors)

From Data to Ficta: A Critical Reflection on Visual Analytics in the Age of Generative Models

Ignacio Pérez-Messina[1], Silvia Miksch[1], and Christian Tominski[2]

[1] TU Wien, Institute of Visual Computing & Human-Centered Technology, Austria
[2] University of Rostock, Institute for Visual & Analytic Computing, Germany

Abstract

Over the last decades, the pervasiveness of data and the rate at which they are produced have propelled the rapid growth of Visual Analytics (VA) methods for interactive analysis and decision-making. This foundation is now shifting as computers increasingly not only process data, but also produce them through generative models. Revisiting the notion of synthetic data, we oppose data to capta (observational inscriptions) and introduce ficta: model-produced outputs that enter VA sessions as an analytic substrate with their own means of interactive production. In ficta-driven workflows, the substrate of analysis is no longer a static representation from which knowledge can be distilled; interaction becomes entangled with what is available to observe, and visualizations come to mediate not a single world but a plurality of model-admissible alternatives. We argue that this breaks VA's classic data-driven epistemic contract and motivates a reevaluation of core assumptions behind knowledge production in the age of generativeness.

CCS Concepts

• ***Human-centered computing*** → ***Visualization theory, concepts and paradigms;***

1. Introduction

Visual Analytics (VA) enables interactive exploration and analysis of large and complex datasets through visual interfaces to support insight generation and decision-making [CT05, KAF*08]. Traditionally, VA is grounded in a data-driven assumption: Analysis begins with samples or observations of the world, and interaction serves to navigate, transform, and interpret this fixed substrate into an informed representation of reality [ALA*18]. This assumption is increasingly losing ground as model-generated outputs (synthetic datasets, simulations, generated training data, and so on) become central to knowledge production so that what is analyzed is often no longer a static substrate but generated during analysis by intertwined human and machine agency. To name and examine this paradigmatic shift, we introduce the concept of *ficta*, and we argue that its rise requires rethinking VA's core principles to embrace not only its challenges but also its possibilities.

In this paper, we present a critical reflection on VA in the age of generative models. As the use of synthetic data for analytical purposes is on the rise, we argue that their generative procedures cannot be ignored and excluded from our understanding of interactive knowledge production. In this context, *ficta* reframes *synthetic data* in a similar way as *capta* [Dru11] provides us with a critical notion of *data*. Based on the critique of the representational model of data [Off24], we develop the concept of *ficta* as *data with a higher degree of freedom* [GL13]. Just as the term *data* rhetorically collapses the methodological particularities of measurement that *capta* make visible, the category of *synthetic data* tends to collapse the artificiality, generative mechanisms, and open-ended multiplicity whose absence would render it just another form of data. Ficta do not imperfectly represent the world, but they stand for a plurality of model-admissible realities whose validity depends on the analytical task rather than on ground truth distance [Off25], shifting knowledge production from interpreting observational traces to interacting with, producing, and deciding over possible worlds during analysis. Our contributions are:

- Sec. 2 introduces *ficta* as a critical category for synthetic data, distinguishing them from *capta* to clarify their different epistemic roles.
- Secs. 3-4 propose the *Simple Generative Visualization Model* as an extension of van Wijk's model, to instantiate how ficta-driven workflows disintegrate VA's data-driven epistemic contract.
- Sec. 5 discusses implications and future research directions for VA in the age of generativeness, outlining how model, interaction, and visualization concepts in VA must adapt.

2. From Data and Capta to Ficta

In this section, we develop the concept of *ficta* as a distinction within data that stands in complement to *capta*.

2.1. Data and Capta

When we talk about data, we refer to a myriad of different things having completely different characteristics: hyperspectral images taken by high-end scientific equipment, responses from people in controlled experiments, tweets, scans of old pictures, geographical coordinates that define boundaries and events over the globe, labels annotated manually or automatically, and so on. What could phenomena of such different origins possibly have in common for us to talk about them indiscriminately as "data"?

In the classical sense of data-as-observations: A datum is an *inscription* produced at one point in time with a strict reference to "external reality". Known as the representational model of data [Off24], this is the view adopted in extant models of VA (e.g., Andrienko et al. [ALA*18]), such as the one shown in Fig. 1**A**. It represents an epistemic stance in which data are conceptually unified by omitting the methodological particularities of their origin [Dru11], resulting in the idea that all data are simply a representation of some "aspect" of an ontologically static reality, where measures and knowledge systems are objective and unchanging: "Data reflect some piece of the reality (real world), which is the subject of analysis." [ALA*18, p.276] It is important that we contrast this epistemic stance with its modern critique. As succinctly put by Austrian philosopher of science Paul Feyerabend: "Science knows no 'bare facts' at all; the 'facts' that enter our knowledge are already viewed in a certain way and are, therefore, essentially ideational." [Fey93, p. 11]

If reality is the subject of analysis, then we shall call data the *substrate of analysis*. Next, we argue that there are two different kinds of what we call data: capta and ficta.

The epistemic stance of data-as-mediation is taken into visualization theory by Johanna Drucker, who coined the term *capta* to challenge the apparently neutral status of data [Dru11]. For Drucker, *data* (Latin for "given") is commonly treated as unmediated access to reality, whereas *capta* (Latin for "captured") emphasizes that what we work with is taken and produced through acts of selection, framing, measurement, and interpretation (see Fig. 1**B**). The term capta includes the idea that when we consider data to be the ground truth, we obscure the fact that they are the product of observation processes that are already biased by the act of observation itself. Capta are produced by historically-situated devices that shape what we see as "reality".

Drucker's permutation of data into capta leaves us with the question of where model-generated outputs, commonly still regarded as data, stand today? A video description from IBM reads:

> "Synthetic data is artificially generated data versus data based on actual events, but it's not 'fake' data. It replicates the properties of real data without the troubles of capturing it, such as confidentiality, low-volume, or expensive-to-validate. With synthetic data, it's easier and less costly to train AI models, however, it's not a panacea. For example, synthetic data may not fully represent the unexpected events that happen in the real world." [IBM23]

This description of synthetic data euphemistically conceals what is interesting here to us: It is not that synthetic data "may not fully represent unexpected events in the real world", but that *synthetic data do not represent events in the real world at all*. The datum-as-inscription loses its reference to an actual reality, and that is the whole point of it. What, then, does it refer to?

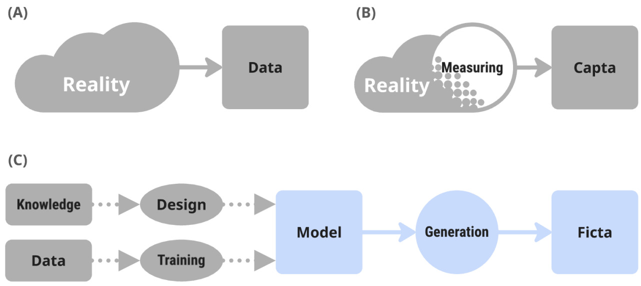

Figure 1: *Two epistemic stances about data and their relation to reality: (A) Data as unmediated access to reality. (B) Capta as the measuring of reality through observation devices. (C) Schematization of the process by which ficta are created.*

By "world," we mean here the conceptual function that defines, according to some rules (e.g., physical or natural laws), the space of what is plausible within it. Synthetic data point to *a* world, otherwise they would not have any value. It is just not *our* world. Synthetic data rather reference many *internally plausible* worlds that are exactly *not* the real world. If we assume there is one actual reality, then "not reality" is not a single opposite state as there is no unique counter-world that stands in symmetric negation to the actual one. Rather, the negation of a specific world opens a space of alternatives: infinitely many worlds that differ from the actual world in small or large ways. Generative models operate in this space: They sample from a distribution of admissible worlds shaped by design, training, and priors.

2.2. Ficta

Building on Drucker's "data are capta" [Dru14, p. 128], we introduce ficta as a concept for all data that are *not* capta, that is, any inscription that is not directly derived from an observation of reality. Like capta, *ficta* (Latin for "fictitious" or "fabricated") are not simply given, but unlike capta, they are not taken from the world by measurement but are procedurally generated by a model. Often referred to under the umbrella term "synthetic data" (see Offenhuber [Off24] for a typology), the term ficta stresses the epistemic-function shift from data-as-evidence to data-as-possibility, helping to release the overburdened concept of data.

Ficta can be defined as data whose time of inscription is completely arbitrary with regard to its reference. Thus, ficta are not necessarily produced by computers, but for our purposes here we will also consider that their inscriptional origin is traced back to a computer model. The term ficta denotes data whose immediate relation to the world is no longer observational but mediated by a generative mechanism: What they refer to is first and foremost the model's learned or specified structure, not a directly measured phenomenon. Fig. 1C roughly schematizes the pipeline by which ficta are created, dividing it into two broad phases: Knowledge/data are used to design/train models, which are then used to generate ficta.

While, strictly speaking, capta are also subject to the problem of synthesis (on account of being ideational objects), and ficta inherit issues of capta (because they are trained on the same historically situated selections, exclusions, and biases), we put the emphasis on ficta where a whole paradigm shift in analysis takes place. Where capta are anchored in acts of measurement, ficta are anchored in acts of creation and therefore in the proliferation of alternatives: They can output not one record of what happened, but many possible versions of what could happen. This multiplicity shifts the epistemic burden from "Did we capture/interpret it correctly?" to "Which world are we taking as plausible?"

As the direct link to the world is severed in ficta (in fact, synthetic data often try to detach from the world to ensure privacy while preserving plausibility), generated data are no longer subject to what has actually been (objective reality). In order to remain plausible, however, ficta must comply with what we believe to be correct models of reality, again introducing the biases that engender capta and are engendered by capta. But even if we accepted one model as the best possible choice, the fact would still remain that ficta are arbitrary and contingent, and constitute only a subset of possibilities out of infinite others. The danger for VA is that being unaware of the distinction between capta and ficta collapses fundamentally different epistemic statuses into a single, misleading category of the "given". Under the common label "data", the different roles that VA must adopt will remain indistinguishable or be outright missed. This consideration is not purely theoretical, but it should reshape the way we design for VA.

3. The VA Epistemic Contract Shift

The epistemic contract of VA is the implicit agreement that an observational dataset provides access to the objective external reality, and that VA systems, through interactive transformations and visual encodings, help analysts explore hypotheses, produce truthful knowledge, and support decisions on the basis of this fixed evidential substrate without changing it. This "data-driven foundation" is reflected in common models of VA that position data as the starting point and insight or knowledge as the end-point. VA provides a feedback loop between these poles through interaction, yet it is still typically conceptualized as a fundamentally unidirectional mapping from data to (new) knowledge [SSS*14]. From an information-theoretic perspective, this mapping can be understood as a progressive compression in which knowledge is treated as latent in data's "uncompressed" form and gradually distilled into recognizable visual patterns and, ultimately, into decisions as entropy is reduced [CJ10].

We claim that this data-driven epistemic contract breaks under generative models and summarize this shift from capta to ficta in the following three postulates:

P1 Artificiality
Ficta are not a function of reality but of a model: Their immediate causal origin is computational generation rather than measurement, so their reference is mediated by the generator's priors, training, and specification. This challenges the representational understanding of data and the common notion of ground truth [Off25].

P2 Generativeness
Ficta are not static but contingent on interaction: The analytic substrate is co-produced during the VA session, as user actions (e.g., prompts, constraints, feedback) directly steer what is generated and thus what becomes available for analysis.

P3 Plurality
Ficta do not point to a single world, but to a *plurality* of model-admissible worlds. Rather than yielding one definitive answer, generative mechanisms typically produce families of alternatives whose differences are not mere noise but constitutive of what the model can express. Consequently, ficta-driven VA cannot be framed as inference over a closed set, but as the discernment, curation, and governance of a multiplicity.

The epistemic shift is not a matter of replacing one better foundation with another but of recognizing that data-driven and ficta-driven workflows constitute two inherently different modes of knowledge production. Data-driven VA treats knowledge as something to be distilled from observational traces in an effort to remain accountable to an external reality through reduction. Ficta-driven VA, by contrast, treats knowledge as something to be constructed from generated alternatives in an effort to govern and exhaust a space of plausible worlds shaped by models. Each supports different modes of understanding, and the contribution of ficta is precisely to expand what VA can know by explicating new theoretical (visual, analytical, and cognitive/perceptual) demands through which VA's models of knowledge production can be revised and further articulated. Next, we take a step in this direction by introducing ficta into one of VA's core models.

4. Simple Model of Generative Visualization

In the age of generative models, capta are not simply replaced by ficta: Ficta enter VA together with their own means of production, enacted through interaction with visual and analytical means during analysis. We therefore propose a *simple generative visualization model* as an extension of van Wijk's fundamental model [vW05], wherein we replace, in Fig. 2, the original *Data* component by a *Ficta* container accompanied by a *Model* container and a *Generation* process, with explicit connections to the rest of the pipeline.

In van Wijk's original model of visualization, the image produced by the visualization at time t is expressed as

$$I(t) = V(D, S, t), \tag{1}$$

where D denotes the (fixed) data and S the specification. In our simple generative visualization model, the static data substrate is replaced by a time-dependent ficta substrate $F(t)$ produced through a generative process. Accordingly, image I becomes

$$I(t) = V\Big(F(t), S_V(t), t\Big), \tag{2}$$

where $S_V(t)$ denotes the visualization specification that may be adapted through interaction. The ficta substrate is generated by a process G that depends on a model M and on a user-steerable generation specification $S_G(t)$, including constraints, prompts, seeds,

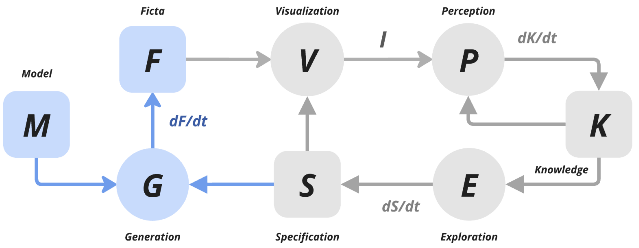

Figure 2: *Extending the simple visualization model with generative capabilities. In blue are the added components of the model, which replace the original Data component. The Model–Generation–Ficta ensemble, together with the change of Ficta dF/dt schematizes a basic generative ficta-driven workflow.*

and parameters. We write this as

$$\frac{dF}{dt} = G\Big(M, S_G(t), t\Big). \tag{3}$$

This formulation articulates that interaction in ficta-driven pipelines can act not only on the visualization (via S_V), but also on the production of the analytical substrate itself (via S_G), thereby coupling visual reasoning with the iterative production of the visual substrate. Note that the training phase and the steps necessary to arrive at the model (and even the parameters and model states) are becoming increasingly irrelevant to users, who only want to gain insight from ficta with an off-the-shelf solution. Therefore, M is defined as an originating, immutable component here.

Replacing van Wijk's single *Data* container with our *Model–Generation–Ficta* triad reframes ficta-driven VA as an analytic session in which the substrate is not assumed from the start, but is iteratively constructed from an initial state $F(t_0)$, which may be void. The *Model* instantiates P1 by providing the space from which the session can draw: an underlying search space whose scope may exceed that of any capta set. The *Generation* process operationalizes P2 : Interactive exploration E no longer merely adjusts a visualization specification, but actively elicits and steers the generation of ficta via $S_G(t)$, such that each decision redefines what counts as the current substrate of analysis. The *Ficta* container reflects P3 : A ficta set is never "complete" as it instantiates only a tiny point in a model's search space, and it can grow unbounded in principle through dF/dt, requiring visual analysis to be highly dynamic and to account for its fabricated and contingent status.

5. Discussion and Research Directions

In his 2002 book [Wol02], Stephen Wolfram announced the advent of a new scientific paradigm brought forward by the study of simple cellular automata that produce complex patterns. While this might be an overestimation of cellular automata, his intuition about how procedural generativeness changes the nature of science might not have been far off. VA has a vantage point on this critical shift taking place under our feet: No other science is called upon to open the black box of algorithms by making them highly visual and interactive [MPG*14] or tasked with designing insight through diagrams of feedback loops [Alb23]. In this paradigmatic shift that will affect all science from its capta-driven foundation, VA will have to understand how this new mode of knowledge production works and create the tools for it. In order to do that, it will have to revisit its own foundations and concepts in terms of models, interaction, and visualization.

5.1. Models

In VA, models have originally been seen as tools for squeezing the knowledge out of data faster [KKEM10], or as objects of inquiry themselves, e.g., in the exploration of model-states [PBM23]. In the ficta regime, however, and following from P1 , the model is no longer merely an instrument applied *to* a dataset: it becomes the condition *of* the dataset, acting as a "reality sandbox" that can present us with more than just one account of events.

These accounts can be largely unstable. Models fabricate realities ultimately using statistical methods, but that does not ensure uniformity in their output. One of the arguments used to show the drawbacks of considering statistics alone is Anscombe's Quartet, a clever arrangement of bivariate (artificial) datasets that exhibit the same statistics and model fits. But visualizing these data reveals highly discrepant patterns. Such effects have been reproduced with other data, including phylogenetic trees [RSVR18]. The argument of going beyond purely statistical methods for data analysis becomes even more relevant when data are being produced by models that statistically mimic capta: We know that qualitatively different worlds and interpretations can arise from the same measures. This has direct implications for validation. Approaches based on ground-truth comparison (e.g., evaluating generated outputs against an assumed reference) remain important in bounded settings [SNW*26], yet they do not fully address the ficta regime, because the concept of ground truth itself is challenged in a post-representational regime. As Offenhuber proposes, the quality of synthetic data and their value should be evaluated "only in relation to a particular purpose" [Off24, p. 7] and not in absolute terms, that is, not in reference to a ground truth.

Much of VA's relationship to AI has so far been framed either as support for model optimization (e.g., labeling, cleaning, and explaining machine learning data [SSSEA20, JCRC24]) or as AI-assisted analysis in which models collaborate alongside humans in the loop (e.g., agentic visualization [DWML*25, SodHB*25]). While valuable, they leave the data-driven contract largely intact, whereas ficta directly destabilize VA's core assumption. What is missing is the VA analog of generative AI: an interactive visual interface that treats generative models not primarily as objects to be trained or interpreted, but as a productive medium for visual knowledge production in a way so far unexplored while developing new methodological frameworks for the validation of their generated analytical substrate.

5.2. Interaction

As a "science of interaction" [PSCO09], VA has struggled to incorporate the idea that the analytical substrate may itself, and not just the interface, be subject to manipulation. In ficta-driven pipelines, the representational model of data becomes actively counterproductive because it masks the fact that the substrate is free from the "necessity" of reality. Turning around one of the statements of semantic interaction "Interaction is itself data." [WDC*18, p. 289], the ficta regime invites the inverse claim: *Data are themselves interaction*. From P2, it follows that once interaction can elicit generation, the analytic substrate no longer needs to pre-exist the session as a dataset, but must be generated on-the-fly.

In VA, interaction has traditionally been framed as a means for analysts to interrogate capta by iteratively re-encoding, filtering, and zooming to correct interpretations, as in the familiar "overview first, details on demand" mantra [Shn03]. Most visualization task taxonomies focus on observations. As the origin of data shifts from measurement to fabrication, task taxonomies can no longer treat analytic work primarily as observations. Brehmer and Munzner's visualization task typology [BM13] showcases this limited understanding of tasks. Most tasks, even when including interactions, ultimately presume a fixed data substrate in which users search, filter, annotate, compare, or present. The exception that confirms the rule is the *produce* task, which covers interactions that change the data or create new artifacts, yet it remains completely under-specified. This reflects how thoroughly VA theory has been oriented toward observation and interpretation, and how little conceptual work has investigated interventions that produce and alter the analytical substrate itself. The current scenario demands that "produce tasks" be expanded into a richer family of operations that articulate how analysts take part in the writing and editing of the analytic sources.

On the other hand, progressive VA has already started to consider what happens when interaction becomes entangled with what is available to see [SPG14, FFS24], how to couple interaction with algorithmic output while keeping responsiveness within interactive thresholds [MPG*14, ASSS18], as well as provenance for model steering and adaptive systems [XOW*20], and persistent interaction through user-generated artefacts [PMCSM25]. However, work in this direction needs to expand beyond fixed data and algorithms and into latent spaces offering endless generative possibilities. Conceptual techniques such as Speculative Execution [SBS*18], which leverage a mixed-initiative approach to guided exploration of machine learning models, incorporate the idea of parallel navigation through possible outcomes at its core. Generative VA systems will need to develop similar engines to produce ficta sets for their inspection and comparison. This kind of exploratory analysis, which goes beyond purely visual, will most probably need guidance-enhanced VA to be coupled with progressive generation [PMAC*25] as single-initiative systems will hardly be a reliable analytical option. This frames a key research direction for VA: designing *interaction-first systems* in which the primary object of analysis is not given data but the evolving ficta corpora that interaction brings into being, as well as task taxonomies that incorporate this new relation to the analytical substrate.

5.3. Visualization

The power of visualization lies in its capacity to phenomenologically collapse inscriptions that originate from distinct phenomena by translating them into familiar, well-studied diagrammatic forms. Through this translation, the outcomes of heterogeneous processes (e.g., biological, social, physical, artificial, and so on) are brought into the same cognitive–perceptual channels. This is productive precisely because it enables pattern recognition and reasoning across domains that would otherwise remain incommensurate (i.e., each domain would need its own encodings, were there any possible encodings at all). Yet therein lies also the heart of its challenge in the age of generativeness: By equating the product of natural and artificial processes, visualization misses and inadvertently disguises the fundamental consequences of generativeness for knowledge production.

Uncertainty in visualization has so far mostly been understood as the difference between data and reality [KDJ*21]. In the classical setting, uncertainty-aware visualization informs a boundary between measurement and inference [SSK*15, p.242]. With ficta, uncertainty is no longer about data quality but rather about *which world we are looking at*. Variability becomes ontological (a family of generated outcomes) rather than epistemic (imperfect access to a single reality). This shift forces visualization to move from decorating evidence with error bars to staging a critical encounter with alternative realities. Should uncertainty, in this case, encode the rarity of a data point (or pattern) within the whole domain of outputs of a generative model (i.e., its contingency)?

Borrowing from Gelman and Loken's critique of "researcher degrees of freedom" and the way exploratory choices can steer results without overt misconduct in "the garden of forking paths" [GL13], we could call ficta "data with a degree of freedom". This is a double-edged sword: It can either double the contingency (findings can depend both on the exploratory path taken through analyses and on which generated world was ultimately examined or reported) or open the possibility of reducing contingency by reasoning over distributions or ensembles (multi-modal, multi-run data) [KH13] of generated sets rather than treating any single sample as decisive (and so escape the inherent contingency of capta; turning P3 into an advantage). To this end, new and alternative visualization methods are needed to reflect the higher complexity of ficta.

6. Conclusion

We have argued that the growing importance of model-generated outputs in VA is not being adequately addressed under the common notions of data and synthetic data. We adopted a critical stance and refined this opposition into *capta* and *ficta*, emphasizing their different epistemic and operational roles. We postulated that artificiality, generativeness, and plurality in ficta introduce a qualitative shift in VA's modus operandi and reflected this in the *Simple Generative Visualization Model* as a minimal extension of van Wijk's model. Finally, we discussed how these categories open challenges for VA: designing interaction-first systems that can start from scratch and build evolving ficta corpora, extending task taxonomies to include substrate manipulation, and developing visualization methods that help reason over ensembles of dynamic and contingent outputs. Together, these contributions aim to intellectually challenge VA and provide a vision for VA in the age of generative models.

Acknowledgments

This work was partially supported by the Austrian Public Employment Service (AMS) and WWTF grant [10.47379/ICT19047].

References

[ALA*18] ANDRIENKO N., LAMMARSCH T., ANDRIENKO G., FUCHS G., KEIM D., MIKSCH S., RIND A.: Viewing visual analytics as model building. In *Computer Graphics Forum* (2018), vol. 37, Wiley Online Library, pp. 275–299. doi:10.1111/cgf.13324. 1, 2

[Alb23] ALBRECHT K. F.: *Insight by de—sign*. doctoral thesis, Universität Potsdam, 2023. doi:10.25932/publishup-57509. 4

[ASSS18] ANGELINI M., SANTUCCI G., SCHUMANN H., SCHULZ H.-J.: A review and characterization of progressive visual analytics. *Informatics 5*, 3 (2018), 31. doi:10.3390/informatics5030031. 5

[BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. *IEEE TVCG 19*, 12 (2013), 2376–2385. doi:10.1109/TVCG.2013.124. 5

[CJ10] CHEN M., JÄNICKE H.: An Information-theoretic Framework for Visualization. *IEEE TVCG 16*, 6 (Nov. 2010), 1206–1215. doi:10.1109/TVCG.2010.132. 3

[CT05] COOK K. A., THOMAS J. J.: *Illuminating the path: The research and development agenda for visual analytics*. Tech. rep., Pacific Northwest National Lab. (PNNL), Richland, WA (United States), 2005. 1

[Dru11] DRUCKER J.: Humanities approaches to graphical display. *Digital Humanities Quarterly 5*, 1 (2011), 1–21. doi:10.63744/r4ysrh7ae534. 1, 2

[Dru14] DRUCKER J.: *Graphesis: Visual Forms of Knowledge Production*. Harvard University Press, Cambridge, MA, 2014. URL: https://www.hup.harvard.edu/books/9780674724938. 2

[DWML*25] DHANOA V., WOLTER A. S., MOLINA LEÓN G., SCHULZ H.-J., ELMQVIST N.: Agentic visualization: Extracting agent-based design patterns from visualization systems. *IEEE Computer Graphics and Applications 45*, 6 (2025), 89–100. doi:10.1109/MCG.2025.3607741. 4

[Fey93] FEYERABEND P.: *Against Method: Outline of an Anarchistic Theory of Knowledge*, 3 ed. Verso, London and New York, 1993. First published by New Left Books (1975); revised edition (Verso, 1988). Also issued with ISBN 0-86091-481-X. URL: https://theanarchistlibrary.org/library/paul-feyerabend-against-method. 2

[FFS24] FEKETE J.-D., FISHER D., SEDLMAIR M.: *Progressive Data Analysis: Roadmap and Research Agenda*. Eurographics Association, 2024. doi:10.2312/pda.20242707. 5

[GL13] GELMAN A., LOKEN E.: The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University 348*, 1-17 (2013), 3. URL: https://sites.stat.columbia.edu/gelman/research/unpublished/forking.pdf. 1, 5

[IBM23] IBM TECHNOLOGY: What is synthetic data? no, it’s not “fake” data. https://www.youtube.com/watch?v=HIusawrGBN4, March 2023. YouTube video. 2

[JCRC24] JIN Y., CARRASCO-REVILLA A., CHEN M.: iGAiVA: Integrated generative ai and visual analytics in a machine learning workflow for text classification. arXiv:2409.15848, doi:10.48550/arXiv.2409.15848. 4

[KAF*08] KEIM D., ANDRIENKO G., FEKETE J.-D., GÖRG C., KOHLHAMMER J., MELANÇON G.: Visual analytics: Definition, process, and challenges. In *Information Visualization*. Springer, 2008, pp. 154–175. doi:10.1007/978-3-540-70956-5_7. 1

[KDJ*21] KAMAL A., DHAKAL P., JAVAID A. Y., DEVABHAKTUNI V. K., KAUR D., ZAIENTZ J., MARINIER R.: Recent advances and challenges in uncertainty visualization: a survey. *Journal of Visualization 24*, 5 (2021), 861–890. doi:10.1007/s12650-021-00755-1. 5

[KH13] KEHRER J., HAUSER H.: Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics 19*, 3 (Mar. 2013), 495–513. doi:10.1109/TVCG.2012.110. 5

[KKEM10] KEIM D., KOHLHAMMER J., ELLIS G., MANSMANN F. (Eds.): *Mastering the information age: solving problems with visual analytics*. Eurographics Association, 2010. doi:10.2312/14803. 4

[MPG*14] MÜHLBACHER T., PIRINGER H., GRATZL S., SEDLMAIR M., STREIT M.: Opening the black box: Strategies for increased user involvement in existing algorithm implementations. *IEEE TVCG 20*, 12 (2014), 1643–1652. doi:10.1109/TVCG.2014.2346578. 4, 5

[Off24] OFFENHUBER D.: Shapes and frictions of synthetic data. *Big Data & Society 11*, 2 (2024), 20539517241249390. doi:10.1177/20539517241249390. 1, 2, 4

[Off25] OFFENHUBER D.: Synthetic data and the shifting ground of truth, 2025. Talk presented at the Society for the Social Studies of Science (4S) 2025 meeting, Seattle, September 3, 2025. doi:10.48550/arXiv.2509.13355. 1, 3

[PBM23] PICCOLOTTO N., BÖGL M., MIKSCH S.: Visual parameter space exploration in time and space. *Computer Graphics Forum 42*, 6 (2023), e14785. doi:10.1111/cgf.14785. 4

[PMAC*25] PÉREZ-MESSINA I., ANGELINI M., CENEDA D., TOMINSKI C., MIKSCH S.: Coupling guidance and progressiveness in visual analytics. *Computer Graphics Forum 44*, 3 (2025), e70115. doi:10.1111/cgf.70115. 5

[PMCSM25] PÉREZ-MESSINA I., CENEDA D., SCHETINGER V., MIKSCH S.: Persistent interaction: A conceptualization of user-generated artefacts in visual analytics. *Computers & Graphics 129* (June 2025), 104232. doi:10.1016/j.cag.2025.104232. 5

[PSCO09] PIKE W. A., STASKO J., CHANG R., O’CONNELL T. A.: The science of interaction. *Information Visualization 8*, 4 (2009), 263–274. doi:10.1057/ivs.2009.22. 4

[RSVR18] REVELL L. J., SCHLIEP K., VALDERRAMA E., RICHARDSON J. E.: Graphs in phylogenetic comparative analysis: Anscombe’s quartet revisited. *Methods in Ecology and Evolution 9*, 10 (2018), 2145–2154. doi:10.1111/2041-210X.13067. 4

[SBS*18] SPERRLE F., BERNARD J., SEDLMAIR M., KEIM D. A., EL-ASSADY M.: Speculative Execution for Guided Visual Analytics. In *Workshop for Machine Learning from User Interaction for Visualization and Analytics at IEEE VIS* (2018), pp. 1–8. doi:10.48550/arXiv.1908.02627. 5

[Shn03] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization*. Elsevier, 2003, pp. 364–371. doi:10.1016/B978-155860915-0/50046-9. 5

[SNW*26] SACHDEVA M., NARAYANAN C., WIEDENKELLER M., SEDLAKOVA J., BERNARD J.: Systematic validation of LLM-generated structured data — a design space and remaining challenges. *Computers & Graphics 135* (2026). doi:10.1016/j.cag.2026.104545. 4

[SodHB*25] STÄHLE T., OP DE HAAR M. J., BOYER S., SEVASTJANOVA R., NARECHANIA A., EL-ASSADY M.: A design space for intelligent agents in mixed-initiative visual analytics. *CoRR abs/2512.23372* (2025). doi:10.48550/ARXIV.2512.23372. 4

[SPG14] STOLPER C. D., PERER A., GOTZ D.: Progressive visual analytics: User-driven visual exploration of in-progress analytics. *IEEE TVCG 20*, 12 (2014), 1653–1662. doi:10.1109/TVCG.2014.2346574. 5

[SSK*15] SACHA D., SENARATNE H., KWON B. C., ELLIS G., KEIM D. A.: The role of uncertainty, awareness, and trust in visual analytics. *IEEE TVCG 22*, 1 (2015), 240–249. doi:10.1109/TVCG.2015.2467591. 5

[SSS*14] SACHA D., STOFFEL A., STOFFEL F., KWON B. C., ELLIS G., KEIM D. A.: Knowledge generation model for visual analytics. *IEEE TVCG 20*, 12 (2014), 1604–1613. doi:10.1109/TVCG.2014.2346481. 3

[SSSEA20] SPINNER T., SCHLEGEL U., SCHÄFER H., EL-ASSADY M.: explAIner: A visual analytics framework for interactive and explainable machine learning. *IEEE Transactions on Visualization and Computer Graphics 26*, 1 (2020), 1064–1074. doi:10.1109/TVCG.2019.2934629. 4

[vW05] VAN WIJK J.: The value of visualization. In *VIS 05. IEEE Visualization, 2005.* (2005), pp. 79–86. doi:10.1109/VISUAL.2005.1532781. 3

[WDC*18] WALL E., DAS S., CHAWLA R., KALIDINDI B., BROWN E. T., ENDERT A.: Podium: Ranking data using mixed-initiative visual analytics. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (Jan. 2018), 288–297. doi:10.1109/TVCG.2017.2745078. 5

[Wol02] WOLFRAM S.: *A New Kind of Science*. Wolfram Media, Inc., Champaign, IL, 2002. URL: https://www.wolframscience.com/nks. 4

[XOW*20] XU K., OTTLEY A., WALCHSHOFER C., STREIT M., CHANG R., WENSKOVITCH J.: Survey on the analysis of user interactions and visualization provenance. *Computer Graphics Forum 39*, 3 (2020), 757–783. doi:10.1111/cgf.14035. 5