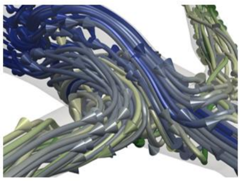
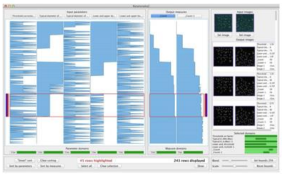
  

The Visualization Landscape for Artificial Intelligence and Data Science

Visualization Interest Group, The Alan Turing Institute, London, UK.

Convenors: Nick Holliman, Patty Holley
Steering group: Hamish Carr, Oliver Davies, Sara Fernstad, Roy Ruddle
Group members: Nick Barlow, Jack Roberts, Ben Bach, Anders Drachen, Daniel Archambault, Cagatay Turkay, Rita Borgo, Alfie Abdul-Rahman, Sarah Dryhurst, Jason Dykes, Elena Simperl, Greg McInerny, Flora Roumpani, Jonathan Roberts, Yannick Wurm.

Call to Action

The opportunities for visualization to contribute to the use and uptake of AI and Data Science are significant, we believe maintaining the UK's international standing in AI and Data Science must include support for visualization research and innovation throughout the TRL scale. The UK can then apply its strengths in rigorous science, design and engineering in this increasingly important field to deliver socio-economic value across a range of themes.

Research Landscape

The AI and Data revolution is generating and exploiting data at rates never before seen in history. Artificial Intelligence and Data Science (AI/DS) provide the technical tools to organise, mine, analyse, categorise, predict and communicate future outcomes using this data. A critical challenge is providing human insight to the operation of, and explanation of results arising from, these complex systems. Visualization provides the key link from the digital to the human, delivering both accurate insights to, and timely cognition of, the meaning of data and outcomes from data-intensive algorithms. Augmentation of human decision-making using data requires the rich two-way link that visualization provides between machines and people [7], a human plus machine mixed intelligence.

Competitors and the Opportunity Landscape

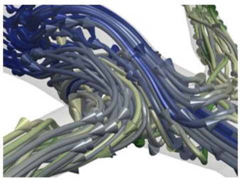
The USA has historically published farsighted reports [1,8] on the future for visualization and has supported research and innovation in the field for some time. This has led to both academic advances and to the creation of leading visualization products such as ESRI ARCGIS, Tableau, QLIC and Microsoft's PowerBI. Commercial US visualization companies have seen a huge increase in revenue in the last few years, for example Tableau's turnover increased more than 30% in the year before they became a billion-dollar company, subsequently acquired by Salesforce for $15.7 billion. Visualization groups exist in major companies e.g. Uber, Netflix & Google.

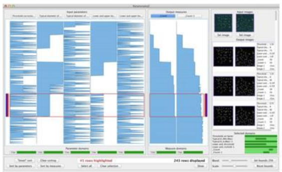
Germany has traditionally invested significantly in visualization research and innovation and has strong research groups within universities and the Fraunhofer network. Tools such as Cubeware and Knime have benefited from this investment, visual analytics has had particularly strong support in academia and industry [9].

Many UK organisations have in-house visualization groups, including the BBC and leading newspapers. Key leaders in business have observed that there is a vital but missing link between data and decision makers [4] that visualization can fill translating the results from AI and Data Science to achieve real impact:
"..success with AI and analytics requires not just data scientists but entire cross-functional, agile teams that include data engineers, data architects, data-visualization experts .."

Who We Are

The Turing special interest group in visualization (#VizTIG) was formed in early 2019 at the Alan Turing Institute and today includes representatives from many of the UK's leading visualization research teams from institutions from within and outside the Alan Turing partner network. The group holds regular monthly meetings and supports events including the highly successful Symposium on Visualization for AI and Data Science in 2019 & 2020, feedback from which has informed this second landscaping document.





Building capacity in skills and knowledge

Data visualization as a specialism covers a broad range of interdisciplinary skills rooted in the science and engineering of computing systems, human perception and cognition, and visual design and communication. In the sciences psychology including vision and visual cognition provide foundational elements, physics and chemistry underpin electronics displays while the mathematical sciences and statistics support algorithmic development. In the arts, humanities and social sciences critique and critical thinking are aligned with ideas from science and technology studies, and art history while the practice of art and graphic design links to the aesthetic design of visualizations and effective visual communications. This leads to the conclusion that the potential range of options in skills provision for visualization is broad.

There is a significant distinction between researchers and practitioners. Researchers are searching for new theoretical approaches to visualization, while practitioners are developing visualizations much like software developers develop code. Broadly researchers aim to discover and validate new principles that impact visualization tools, while practitioners apply these tools to deliver visualizations for end users. This is reflected in the training differences between PhD educated researchers and Masters level educated practitioners. Key skill areas that underpin the highly interdisciplinary visualization community include:

Computing

- Algorithms and data structures
- Software engineering and delivery
- Engineering of data systems
- Machine Learning and AI
- Graphics and interactive systems
- Parallel and HPC systems

Mathematics and Statistics

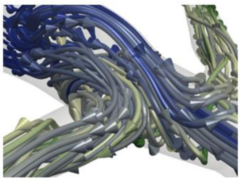
- Linear algebra and geometry
- Calculus and time series
- Descriptive and inferential statistics
- Prediction methods
- Exploratory data analytics
- Visual analytics skills
- Mathematical modelling

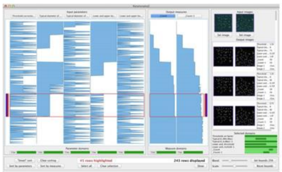
Science, Engineering, Medicine

- Vision and visual cognition
- Behaviour & decision making
- Human centered design
- Human participant studies
- Displays & Interfaces incl. VR & AR.
- Computational science
- Medical data visualization

Arts, Humanities & Social Sciences

- Aesthetics and Semiotics
- Visual and interaction design & art
- Visual narratives & reasoning
- Spatial analysis and geographic data
- Truth, trust and factfulness
- Legal frameworks and responsibilities
- Policy research and decision making





Grand Challenges

While much press coverage of AI and Data Science might lead the reader to the conclusion that all human tasks will become fully automated by computers, it seems much more likely that these new computational methods will most often be used to augment human comprehension and decision-making abilities. In the domain of mixed intelligence (MI) the connection between the machine and the human user becomes critical with the visual forming a well characterised, high bandwidth link between the two [5].

Visualization in a whole host of forms has been shown to be effective in a wide range of AI and Data Science contexts. The audiences for visualizations vary greatly and range from: Data Engineers clarifying and designing data capture and ETL pipelines, Data Scientists learning about the operation of, and debugging, analytics and machine learning systems, Mathematical Modellers exploring the complexities of model inputs and outputs to steer models and inform insights Business Analysts seeking technical insights to data streams, high level Decision Makers taking accurate, timely summaries of complex processes and the Public looking for advice and insight relevant to everyday life. An explosion of interest in data visualizations has been part of the COVID-19 pandemic.

Recent survey work has identified unmet requirements for visualization systems in AI and Data Science [2,3,6], leading to open challenges where there is significant scope for new research and innovation. We expand on these to suggest the following Grand Challenges for visualization research:

1. Examine and prepare data and data collections, e.g. identify missing data, represent data quality.
2. Represent the model training process for comprehension, debugging and verification leading to explainable DS and AI.
3. Compare the quality of resultant models, demonstrate saliency links between outputs and inputs.
4. Explain the choices underpinning, and uncertainty of model outcomes, reduce user misperceptions.
5. Recommendations to leverage new types of display hardware for everyday engagement with data and models.
6. Improved user interaction with data sets, models and model outcomes to support cognition.
7. Scalable visualization methods allowing interrogation at multiple spatial, temporal and security levels.
8. Optimising narrative approaches to presenting data stories for comprehension and decision making.
9. Represent uncertainty visually to communicate data limitations transparently and support decision makers.
10. Rigorous experimental methodologies to show the effects of visualization in controlled and applied settings.
11. Expand the theoretical base for visualization linked to the empirical evidence from experiments.
12. Encourage deep interdisciplinary links across all areas of endeavour underpinning the field including ethics.

Many of these challenges come from the scale and speed of automation and matching that to the bounds of human comprehension, something that is a constantly moving target.

References

[1] Chris Johnson, Robert Moorhead, Tamara Munzner, Hanspeter Pfister, Penny Rheingans, and Terry S. Yoo, NIH/NSF, Visualization Research Challenges, 2005.

[2] Von Landesberger, T., Fellner, D. W., & Ruddle, R. A. Visualization system requirements for data processing pipeline design and optimization. IEEE Transactions on Visualization and Computer Graphics, 2017.

[3] Sacha, Dominik; Kraus, Matthias; Keim, Daniel A.; Chen, Min , VIS4ML: An Ontology for Visual Analytics Assisted Machine Learning, IEEE Transactions on Visualization and Computer Graphics, 2018.

[4] Nicolaus Henke, Jordan Levine, Paul McInerney, You Don't Have to Be a Data Scientist to Fill This Must-Have Analytics Role, HBR, Feb, 2018.

[5] Tamara Munzner, Visualization Analysis and Design, AK Peters, 2014.

[6] Fred Hohman, Minsuk Kahng, Robert Pienta, Duen Horng Chau, Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers, IEEE TVCG, 2018.

[7] Exploratory Data Analysis, John Tukey, Addison Wesley, 1977.

[8] Visualization in Scientific Computing, McCormick, DeFanti and Brown, Computer Graphics, Vol 21, 1987.

[9] Keim, D., Kohlhammer, J., Ellis, G., & Mansmann, F., Mastering the information age: solving problems with visual analytics, 2010.