Zhu-Tian Chen | Research Statement

My research in **data visualization** aims to **advance human-data interaction in augmented reality (AR) environments.** AR is attracting billions of investments from industry (e.g., Metaverse [1]) and government (e.g., Department of Defense [2]) for its ability to revolutionize the way we access data in everyday activities. Yet, while AR hardware and infrastructures are increasingly available, users are still lacking software methods to work with data effectively in AR. Solving this challenge is the key to unleashing the capabilities of AR and enabling new applications in numerous domains such as medicine, urban design, broadcasting, manufacturing, and even daily activities.

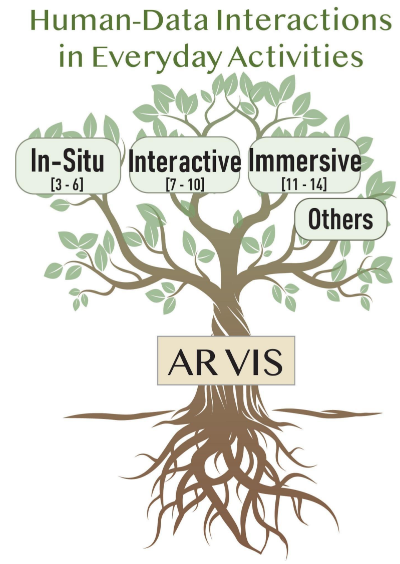

Fig.1 I develop a set of human-AI systems to support the creation and interaction of AR visualizations, enabling users to work with data in in-situ, interactive, or immersive manners in their everyday activities.

In my prior research, I invented a set of **hybrid human-AI systems that enable users to visualize and interact with data in AR environments with low effort and high controllability**. These systems leverage AR visualizations to change how we work with data in everyday activities: They help users perform **in-situ** analyses of data that was previously only accessible offline [3]–[6], conduct **interactive** explorations of data that was originally static [7]–[10], and use **immersive** 3D space to browse data that is challenging to display on 2D screens [11]–[14]. My research has led to 22 publications in top venues, including IEEE VIS, ACM CHI, and TVCG. I **received 3 best paper nominations** in IEEE VIS, the premier visualization conference. The impact of my work extends beyond academia: my works have been accepted to the MIT Sloan Sports Analytics Conference [15], the **premier sports industry conference**, for their commercialization potential in broadcasting, and have been used by and sparked **collaborations with leading tech companies** (e.g., Microsoft, Adobe, T-Mobile) **and elite athletes** (e.g., China National Table Tennis Team, German Soccer Association, U.S. Badminton Olympians).

In the future, I am eager to continue pushing the frontier of human-data interaction in AR by investigating its **technical foundation** and **social impact**, and eventually **enhance the ability of humans to work with data in their everyday activities.**

AR Sports: Support In-Situ Visualization in Physical Environments

One of AR's most compelling features is its ability to visualize data in physical environments. Such an *in-situ* visualization can benefit a variety of tasks that require data in the physical context. Yet, creating in-situ visualizations is challenging since existing methods often require proficient AR programming skills. I developed human-AI systems that **allow users to express their high-level data needs through natural interactions (**e.g., touch, speech, or gaze) **and automatically finish the low-level visualization tasks.** To assess my systems, I use **sports videos** as a reproducible testbed, which involves tremendous data that users (e.g., coaches, fans) want to understand in physical contexts.

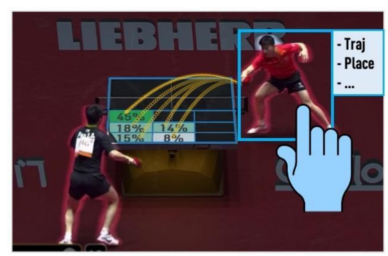

Fig.2 VisCommentator allows users to interact with the video objects directly to select the data to be visualized in the physical scene.

Visualizing data in situ by direct manipulation and natural language

Visualizing sports data in physical contexts is challenging but can help audiences better understand and engage with a game. My observation is that audiences often think of the data in terms of objects or events, such as the *players*' positions and the *ball*'s trajectory following a *shot*. I thus proposed to ask users to express their data needs by describing relevant objects or events, and then use AI models to map them to AR visualizations.

I developed VisCommentator [3] and Sporthesia [4] to support object- and event-based creation, respectively. For object-based creation, VisCommentator uses computer vision (CV) models to allow users to interact with video objects (e.g., players) to select the data of interest (e.g., trajectories). It then automatically visualizes the selected data in the scene (Fig.2) based on a design space of AR sports visualizations. For event-based creation, Sporthesia uses natural language processing (NLP) models to enable users to visualize data in the scene (Fig.3) by describing events (e.g., *"Federer is hitting")* of interest.

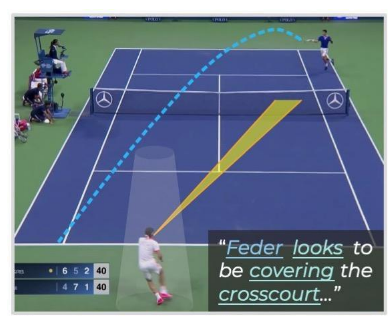

Fig.3 Sporthesia supports users to create AR visualizations by directly describing the events of interest.

I evaluated VisCommentator and Sporthesia with sports spectators and found that both systems allow the users to visualize data in sports scenes with significantly less effort. Both systems open a wide range of applications and have made practical impacts (e.g., VisCommentator was **used by the China National Table Tennis Team to prepare for the Tokyo 2020 Summer Olympics,** Sporthesia was **accepted to the MIT Sloan Sports Analytics Conference for its impact on sportscasting**). VisCommentator also **received an honorable mention award at IEEE VIS 2021** [3].

Adapting visualizations in situ based on physical context and users' gaze

VisCommentator and Sporthesia enable users to visualize data in racket-based sports effortlessly. However, they may fall short in supporting more complex scenarios, such as team sports. To this end, I proposed to automatically visualize data of important players based on the game context and adapt the visualizations based on the users' gaze, which is a fast, rich, and implicit information source that reflects users' intentions.

To ground this idea in practical scenarios, I used AR visualizations to improve fans' experience in watching live basketball videos. I first conducted a user-centered design study [5] to understand basketball fans' in-game data needs and derived a design framework that guides the design of AR visualizations in live basketball games. Based on the findings, I developed iBall [6], a game-watching system for watching live basketball videos (Fig.4). iBall is built based on a tailored CV pipeline, a group of basketball metrics suggested by die-hard fans in a formative study, and gaze-moderated visualizations.

A user study with fans revealed that iBall could help fans better appreciate basketball games without being overwhelmed. The visual designs and interactions in iBall can be naturally extended to further scenarios as AR devices increasingly support eye tracking. As far as I know, I am **the first to use gaze to adapt AR visualizations**. iBall was **accepted by the MIT Sloan Sports Analytics Conference for its impact on sportscasting**, and its design study **received an honorable mention award at IEEE VIS 2022** [5].

Fig.4 iBall automatically visualizes the in-game data of important players (Top) and adapts the visualizations based on the context and user's gaze (Bottom).

Augmented Paper: Enable Interactive Content on Physical Objects

AR can bring interactivity to everyday objects. Applying this feature to paper can make printed content interactive, thereby benefiting a wide range of application scenarios, such as education, office work, and mass media. Yet, it remains unclear how users can convert printed content into interactive without coding and interact with the printed content. I develop human-AI systems to **allow users to convert printed visualizations to interactive through interactive reverse engineering and interact with the printed visualizations through tangible interactions provided by the paper.**

Augmenting printed visualizations through interactive reverse engineering

Imagine a user reading a timeline on a poster with an AR viewer (e.g., a mobile device). To make the timeline interactive, the AR viewer must first understand it, which can be achieved by CV models. But the CV models could fail and don't know how to augment the visualization correctly. I thus developed an interactive reverse engineering method that asks users to steer the machine in understanding and augmenting printed visualizations.

Specifically, I developed MARVisT [7], a mobile authoring tool that allows users to extract visual encodings and elements from printed visualizations, reuse them to visualize new data through simple drag-and-drop interactions, and select how the new visualizations augment the printed ones (Fig.5). I built MARVisT based on a deep learning-based method [9] that interprets printed visualizations and a structured design space [8] that summarizes four ways to augment printed visualizations in AR.

Fig.5 When viewing a timeline on a wall through MARVisT, the user can scan the timeline, reuse its visual encodings and elements to create a new one, and select how to augment it.

My method can be **applied not just to printed but to any static visualizations**, such as those stored in bitmaps or shown in public displays. I **released the code, models, and dataset of my method**. Given its broad applicability, it has received significant attention in the research of chart interpretation, and **has been used directly by many following works** (e.g., [16], [17]) **and deployed internally at Microsoft**.

Interacting with printed visualizations through tangible interactions

Interacting with printed visualizations is challenging since they are presented on paper, a rollable display that is completely different from mainstream displays such as desktop monitors and tablets. Instead of asking the users to use interactions designed for touchscreens (e.g., touch, point), I proposed leveraging the rich tangible interactions provided by paper, such as folding, tilting, or stacking. These natural interactions allow users to manipulate the paper directly to explore printed visualizations.

To realize this idea, I systematically explored the possible mappings through workshops with visualization and HCI researchers, soliciting a structured design space that outlines 81 useful mappings between paper interactions and visualization commands (e.g., pan, filter, zoom), among which 11 were preferred. I then implemented these 11 mappings in an AR system on HoloLens 2 based on gesture recognition models.

In a controlled study, I tested these 11 mappings and found that most of them are intuitive, effective, and engaging. My work is **the first to explore tangible interactions in AR for visualization interactions**. It also gains insights into designing interactions for rollable displays. **This work received an honorable mention award at IEEE VIS 2022** [10].

Fig.6 Users can directly manipulate the paper to interact with the charts.

Immersive Analytics: Use Infinite 3D Spaces for Data Exploration

Unlike small, 2D displays in desktop computers, AR can display digital information in an infinite 3D space and immerse users in the data. I exploit this feature to **develop systems that enable users to experience the data in a first-person perspective and leverage spatial cognition to explore the data in immersive spaces.**

To better explore this idea, I simulated immersive spaces using virtual reality devices as they support a larger field of view than current AR devices. I developed multiple immersive analytics systems for exploring 3D data that is hard to display or interact with on desktop computers. These systems feature novel immersive visualizations, enabling urban planners to select locations for new buildings [14], sports analysts to analyze badminton trajectories [11][12], and archaeologists to reassemble the fragments of artifacts [13]. In the user studies, the experts confirmed that the interactive immersive visualizations allow them not only to see but also to "*feel*" the data. For example, in ShuttleSpace [11], users can see the shuttle trajectories in a first-person view and select trajectories by performing the corresponding stroke, which significantly enhances their understanding of the player's in-game decisions. Deeply rooted in the practical problems of domain experts, my work has made fruitful contributions to academia and real-world applications. Recently, I have **worked with U.S. badminton Olympians and researchers from leading tech companies** (e.g., Adobe) to continue exploring immersive spaces for 3D data exploration.

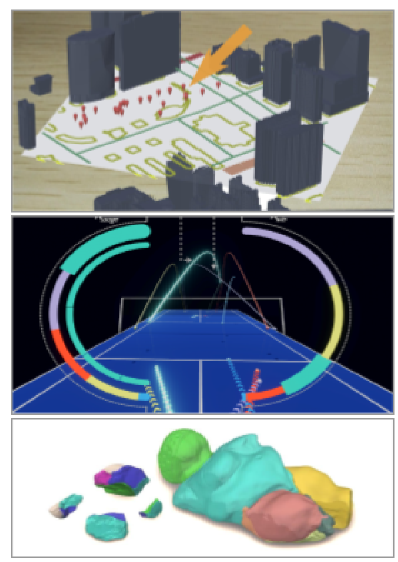

Fig.7 Experts use immersive spaces to explore 3D urban data, badminton trajectories, and fractured artifacts.

Others: Visual Analytics Systems for Data Science

Besides AR visualizations, I also develop interactive visualizations for desktops. The visual analytics systems I built help users monitor streaming data (e.g., Twitter [18]), open the black box of AI models [19], interact with point clouds data [20], and communicate insights through data documents [21]. These systems help domain experts and data scientists solve real-world problems, and meanwhile serve as petri dishes to explore the technical foundations for AR visualizations. For example, the text-data binding technique I developed in CrossData [21] is the base of Sporthesia [4].

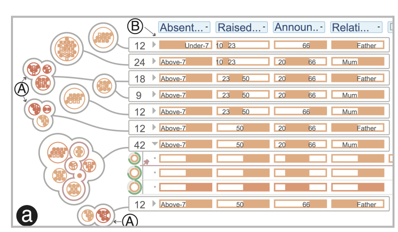

Fig.8 DiscriLens helps users understand the discrimination in ML models.

Research Agenda

My long-term research vision is a future where **anyone can use data anytime and anywhere to support their work and life**. To achieve this vision, I am eager to collaborate with future colleagues on the research problems of human-data interaction in AR in the **technical** and **human** aspects, from the **individual** to **population** levels.

Technical foundations for broader AR visualizations

I have developed technical and design foundations for AR visualizations in my prior systems [3]–[14]. While my systems can work well in the lab, there are still many open issues that must be addressed to fully support AR visualizations in real-world settings. Some opportunities that I am excited to pursue include:

- **Generative power for AR visualizations.** My prior research has strived to enable users to visualize data in AR in situ without coding. However, these methods are built based on manually crafted design spaces. In the dynamic physical world, we need more powerful generative methods to help users create visualizations conditioned by the data, environments, and tasks. I am starting to tackle this issue by exploring **Foundation Models** (e.g., GPT-3, DALL-E).
- **AR visualizations with adaptive appearance**. The physical world is extremely diverse and dynamic. AR visualizations must automatically adapt their visual appearances to the environments (e.g., the background color and ambient light) to ensure users' perceptions. Addressing this fundamental issue can have a huge impact on any AR visual user interface. I consider **Neural Rendering** [22] as a potential solution, which combines CV, computer graphics, and visualization.
- **AR visualizations with proactive behaviors**. Asking users to manually control AR visualizations in a dynamic environment is infeasible. Thus, the visualizations must proactively observe the environment and behave appropriately, such as moving to a position to avoid occlusion or showing a suitable level of detail to avoid overwhelming users. I argue that AR visualizations are similar to intelligent agents in robotics but without physical bodies**.** I have explored using **Reinforcement Learning** to make AR labels proactively adjust their positions in dynamic scenarios [23]. I will continue exploring this cross-disciplinary direction.
- **Reproducible AR testbeds for the community**. Achieving my long-term vision will need the endeavor of a whole community spanning academia and industry. Yet, an innate challenge of AR research is that the physical environments are difficult to distribute and reproduce. My current research uses *videos* as a reproducible testbed to approximate physical environments. In the future, I aim to develop more high-fidelity testbeds that enable others to follow existing research and make progress easily. I have started to develop such a testbed by using **3D Reconstructions** on videos [24].

Human-centered research in AR visualizations

I have consistently worked on the human-aspect issues of AR systems by studying human-data interactions in AR. Building on my prior works, I would like to expand to other human-aspect issues including:

- **Human-AI collaboration with AR visualizations.** My prior works have shown that combining the capability of humans and AI is a key to supporting human-data interaction in AR. However, AI models can make mistakes and lead to severe issues, especially in high-stake AR applications, e.g., surgery, emergency service, and field exploration. I will investigate using **Uncertainty Visualization and Visual Guidance** to help users identify and interactively fix the errors of AI models in AR.
- **Human-human communication with AR visualizations**. I have used users' gaze to customize the AR visualizations in iBall [6]. Yet, such customization can lead to difficulties in human-human communications as different users will see different AR content. *How can users communicate with each other if the physical world looks different to each other?* Moreover, *how can users with AR devices communicate with others without?* I would like to address such an asymmetrical communication issue in the future.
- **Social impact of AR systems.** Ultimately, to democratize AR systems to everyone in the future, I believe we must study the social impacts of AR systems, including but not limited to accessibility, accountability, fairness, privacy, and ethics. I am interested in investigating these computational social science issues that we must face in the AR future.

In summary, I am a data visualization researcher who seeks to facilitate human-data interactions in AR. I will expand my research scope to include more technical and societal factors. I look forward to collaborating with researchers from, e.g., Human-Computer Interaction more broadly, Computer Graphics, Robotics, Machine Learning, Data Science, and Social Science. I am also eager to expand my research to broader applications, extending from sports and official space to hospitals, factories, farms, and beyond, where people need data support beyond the desktop.

References

[1] "Metaverse may be $800 billion market, next tech platform Insights," *Bloomberg Professional Services*, 2021. Available: https://www.bloomberg.com/professional/blog/metaverse-may-be-800-billion-market-next-tech-platform/

[2] J. Novet, "Microsoft wins U.S. Army contract for augmented reality headsets, worth up to $21.9 billion over 10 years," *CNBC*, 2022. Available: https://www.cnbc.com/2021/03/31/microsoft-wins-contract-to-make-modified-hololens-for-us-army.html.

[3] **Z. Chen**, S. Ye, X. Chu, H. Xia, H. Zhang, H. Qu, and Y. Wu, "Augmenting Sports Videos with VisCommentator," *IEEE Trans. Vis. Comput. Graph.,* 2022.

[4] **Z. Chen**, Q. Yang, X. Xie, J. Beyer, H. Xia, Y. Wu, and H. Pfister, "Sporthesia: Augmenting Sports Videos Using Natural Language," *IEEE Trans. Vis. Comput. Graph.,* 2022.

[5] T. Lin, **Z. Chen**, Y. Yang, D. Chiappalupi, J. Beyer, and H. Pfister, "The Quest for Omnioculars: Embedded Visualization for Augmenting Basketball Game Viewing Experiences," *IEEE Trans. Vis. Comput. Graph.*, 2022.

[6] **Z. Chen**, Q. Yang, J. Shan, T. Lin, J. Beyer, X. Haijun, and P. Hanspeter, "iBall: Augmenting Basketball Videos with Gaze-moderated Embedded Visualizations." – submitted to ACM CHI 2023.

[7] **Z. Chen**, Y. Su, Y. Wang, Q. Wang, H. Qu, and Y. Wu, "MARVisT: Authoring Glyph-Based Visualization in Mobile Augmented Reality," *IEEE Trans. Vis. Comput. Graph.*, 2020.

[8] **Z. Chen**, W. Tong, Q. Wang, B. Bach, and H. Qu, "Augmenting Static Visualizations with PapARVis Designer," in *Proc. of ACM CHI*, 2020.

[9] **Z. Chen**, Y. Wang, Q. Wang, Y. Wang, and H. Qu, "Towards Automated Infographic Design: Deep Learning-based Auto-Extraction of Extensible Timeline," *IEEE Trans. Vis. Comput. Graph.*, 2020.

[10] W. Tong, **Z. Chen**, M. Xia, L. Y.-H. Lo, L. Yuan, B. Bach, and H. Qu, "Exploring Interactions with Printed Data Visualizations in Augmented Reality," *IEEE Trans. Vis. Comput. Graph.*, 2022.

[11] S. Ye, **Z. Chen**, X. Chu, Y. Wang, S. Fu, L. Shen, K. Zhou, and Y. Wu, "ShuttleSpace: Exploring and Analyzing Movement Trajectory in Immersive Visualization," *IEEE Trans. Vis. Comput. Graph.*, 2021.

[12] X. Chu, X. Xie, S. Ye, H. Lu, H. Xiao, Z. Yuan, **Z. Chen**, H. Zhang, and Y. Wu, "TIVEE: Visual Exploration and Explanation of Badminton Tactics in Immersive Visualizations," *IEEE Trans. Vis. Comput. Graph.*, 2022.

[13] S. Ye, **Z. Chen**, X. Chu, K. Li, J. Luo, Y. Li, G. Geng, and Y. Wu, "PuzzleFixer: A Visual Reassembly System for Immersive Fragments Restoration," *IEEE Trans. Vis. Comput. Graph.*, 2022.

[14] **Z. Chen**, Y. Wang, T. Sun, X. Gao, W. Chen, Z. Pan, H. Qu, and Y. Wu, "Exploring the design space of immersive urban analytics," *Vis. Inform.*, 2017.

[15] "MIT Sloan Sports Analytics Conference - Sports Analytics, Business, & Technology." https://www.sloansportsconference.com.

[16] A. Tyagi, J. Zhao, P. Patel, S. Khurana, and K. Mueller, "Infographics Wizard: Flexible Infographics Authoring and Design Exploration," *Comput. Graph. Forum*, 2022.

[17] J. Lin, Y. Cai, X. Wu, and J. Lu, "Graph-Based Information Block Detection in Infographic with Gestalt Organization Principles," *IEEE Trans. Vis. Comput. Graph.*, 2021.

[18] Y. Wu, **Z. Chen**, G. Sun, X. Xie, N. Cao, S. Liu, and W. Cui, "StreamExplorer: A Multi-Stage System for Visually Exploring Events in Social Streams," *IEEE Trans. Vis. Comput. Graph.*, 2018.

[19] Q. Wang, Z. Xu, **Z. Chen**, Y. Wang, S. Liu, and H. Qu, "Visual analysis of discrimination in machine learning," *IEEE Trans. Vis. Comput. Graph.*, 2021.

[20] **Z. Chen**, W. Zeng, Z. Yang, L. Yu, C.-W. Fu, and H. Qu, "LassoNet: Deep Lasso-Selection of 3D Point Clouds," *IEEE Trans. Vis. Comput. Graph.*, 2020.

[21] **Z. Chen** and H. Xia, "CrossData: Leveraging Text-Data Connections for Authoring Data Documents," in *Proc. of ACM CHI,* 2022.

[22] A. Tewari, O. Fried, J. Thies, V. Sitzmann, S. Lombardi, K. Sunkavalli, R. Martin-Brualla, T. Simon, J. Saragih, M. Nießner, R. Pandey, S. Fanello, G. Wetzstein, J.-Y. Zhu, C. Theobalt, M. Agrawala, E. Shechtman, D. B. Goldman, and M. Zollhöfer, "State of the Art on Neural Rendering," *Comput. Graph. Forum*, 2020,

[23] **Z. Chen**, D. Chiappalupi, T. Lin, Y. Yang, J. Beyer, and H. Pfister, "Reinforcement Learning-Based View Management for AR labels in Dynamic Environments," in preparation.

[24] T. Lin, **Z. Chen**, A. Aouididi, P. Hanspeter, and W. Jui-Hsien, "Reconstructing 3D badminton games from single camera videos," in preparation.