

Intent Tagging: Exploring Micro-Prompting Interactions for Supporting Granular Human-GenAI Co-Creation Workflows

Frederic Gmeiner*
Microsoft Research
Redmond, WA, USA
Carnegie Mellon University
Pittsburgh, PA, USA
gmeiner@cmu.edu

Nicolai Marquardt
Microsoft Research
Redmond, WA, USA
nicmarquardt@microsoft.com

Michael Bentley
Microsoft
Redmond, WA, USA
mbentley@microsoft.com

Hugo Romat
Microsoft
Redmond, WA, USA
romathugo@microsoft.com

Michel Pahud
Microsoft Research
Redmond, WA, USA
mpahud@microsoft.com

David Brown
Microsoft Research
Redmond, WA, USA
dabrown@microsoft.com

Asta Roseway
Microsoft Research
Redmond, WA, USA
astar@microsoft.com

Nikolas Martelaro
Carnegie Mellon University
Pittsburgh, PA, USA
nikmart@cmu.edu

Kenneth Holstein
Carnegie Mellon University
Pittsburgh, PA, USA
kjholste@cs.cmu.edu

Ken Hinckley
Microsoft Research
Redmond, WA, USA
kenneth.p.hinckley@gmail.com

Nathalie Riche
Microsoft Research
Redmond, WA, USA
nath@microsoft.com

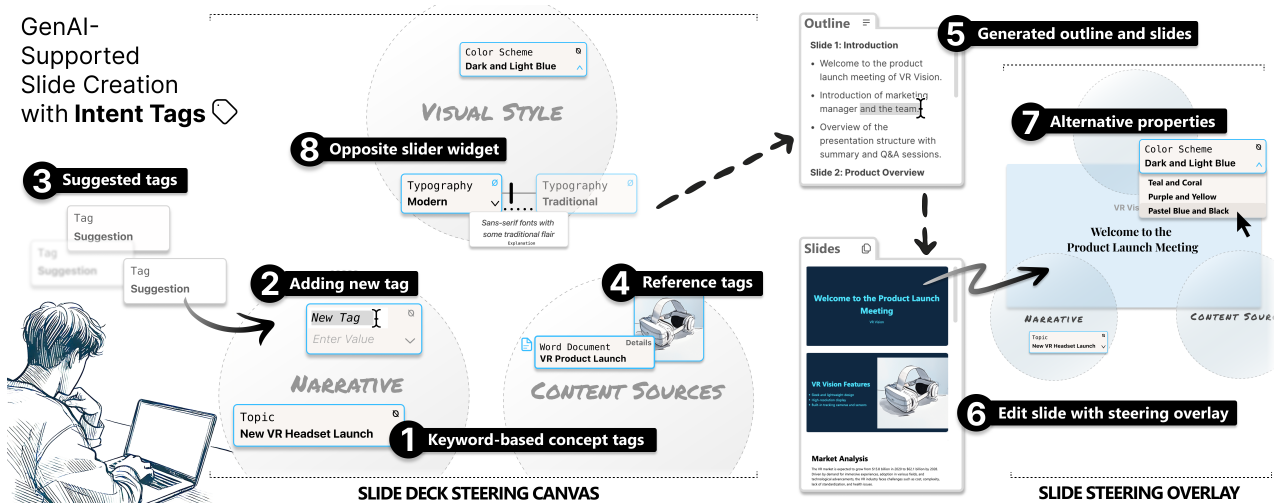


Figure 1: IntentTagger, an intent tag-based LLM-driven slide creation system: Users can steer slide generation with (1) keyword-based concept tags by (2) adding them directly, (3) choosing from suggestions, or (4) adding existing content. (5) A user-editable outline and slide deck are generated grounded in all selected tags. (6) For fine-grained changes, the user can change individual slides with the same (scoped) intent tagging mechanism. Each tag allows exploring alternative properties in a dynamically generated (7) drop-down list, and tag values can be fine-tuned through the (8) opposite slider widget.

*Work done as an intern researcher at Microsoft Research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
CHI '25, April 26-May 1, 2025, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1394-1/25/04
<https://doi.org/10.1145/3706598.3713861>

Abstract

Despite Generative AI (GenAI) systems' potential for enhancing content creation, users often struggle to effectively integrate GenAI into their creative workflows. Core challenges include misalignment of AI-generated content with user intentions (intent elicitation and alignment), user uncertainty around how to best communicate their intents to the AI system (prompt formulation), and insufficient flexibility of AI systems to support diverse creative workflows (workflow flexibility). Motivated by these challenges, we created IntentTagger: a system for slide creation based on the notion of

Intent Tags—small, atomic conceptual units that encapsulate user intent—for exploring granular and non-linear micro-prompting interactions for Human-GenAI co-creation workflows. Our user study with 12 participants provides insights into the value of flexibly expressing intent across varying levels of ambiguity, meta-intent elicitation, and the benefits and challenges of intent tag-driven workflows. We conclude by discussing the broader implications of our findings and design considerations for GenAI-supported content creation workflows.

CCS Concepts

• **Human-centered computing** → **Interaction paradigms**.

Keywords

intent tagging, human-AI interaction, human-AI co-creation, generative AI, rich content creation

ACM Reference Format:

Frederic Gmeiner, Nicolai Marquardt, Michael Bentley, Hugo Romat, Michel Pahud, David Brown, Asta Roseway, Nikolas Martelaro, Kenneth Holstein, Ken Hinckley, and Nathalie Riche. 2025. Intent Tagging: Exploring Micro-Prompting Interactions for Supporting Granular Human-GenAI Co-Creation Workflows. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26–May 1, 2025, Yokohama, Japan. ACM, New York, NY, USA, 31 pages. <https://doi.org/10.1145/3706598.3713861>

1 Introduction

Generative AI (GenAI) models have become increasingly powerful in content creation tasks, and are rapidly being integrated into a range of professional applications, including coding IDEs [16, 32], text processors [13, 34], image and video editors [1, 35, 73], and office suites [58]. These developments offer tremendous promise for AI as a co-creation tool steered by people, but significant challenges remain:

- (1) It is hard to **align AI-generated content with user intentions** (*intent elicitation and alignment*) [16, 88], often due to the difficulty in balancing the flexibility of free-form text prompts with guided interfaces;
- (2) Users often struggle with **formulating effective prompts and understanding an AI system’s capabilities**, leading to trial-and-error interactions (*prompt formulation*) [49, 84, 100];
- (3) The iterative nature of content creation, which requires continuous reflection and refinement, further complicates intent elicitation and prompt formulation since **users’ needs and intents aren’t often completely clear upfront** [14, 79, 82, 89];
- (4) Many **GenAI systems impose workflows** that force users to adapt their commonly used creative processes or generate content with each refinement, further complicating human-AI collaboration (*workflow flexibility*) [75, 87].

In this work, we explore potential solutions to these challenges around *intent elicitation and alignment*, *prompt formulation*, and *workflow flexibility* through the lens of GenAI-supported rich content creation tasks, such as slide deck creation. Slide deck creation involves complex decisions around content structure, visual style, media content, slide sequencing, and narrative flow, all influenced

by implicit factors like audience or presentation duration. Often, slide decks reference external documents, evolve over time, and may be duplicated and changed for different audiences and purposes.

Technically, GenAI multi-modal foundation models offer powerful capabilities to support such complex content creation tasks, for example, by transforming documents or synthesizing various content sources into a single slide deck [21, 30, 95]. However, most human-GenAI interaction challenges are amplified in the context of such complex content creation scenarios—the diverse and non-linear workflows in slide deck creation, coupled with the difficulty of expressing requirements upfront, make it challenging for users to effectively integrate GenAI into their specific creation process.

To address these challenges and empower users in rich AI-assisted content creation tasks, we propose **intent tags** as a granular and flexible technique for GenAI-supported workflows via graphical micro-prompting. Intent tags represent atomic conceptual units that encapsulate a single aspect of a user’s intent. Intent tags help users steer content generation in non-linear workflows—the user instantiates multiple intent tags and edits each individually through adaptive UI elements that allow for granular control of content generation and transparent intent elicitation. To support a creative design process where users rapidly explore multiple alternatives, iterate, and reflect on the outcomes, our system proactively surfaces AI-generated suggestions through intent tags. These suggestions appear in an unobtrusive manner, allowing people to see the parameter spectrum [89] of possible outcomes that they might not have considered or been aware of and use them to iteratively refine the system’s outputs as desired.

To explore further possibilities of intent tagging for human-GenAI co-creation workflows for rich content creation, we created **IntentTagger**: a GenAI-driven system allowing users to iteratively create and modify slide deck presentations using intent tags on a 2D canvas interface. Through intent tags, users steer the generation of single slides or the entire deck by defining keywords (*concept tags*) or including content from other documents or images (*reference tags*). Under the hood, IntentTagger utilizes an LLM for slide generation. IntentTagger also dynamically generates adaptive UI elements such as context-related *tag suggestions*, dynamic *drop-down lists* with *slide preview tooltips*, and interactive *slider widgets* to help users fine-tune tag expressions.

We explore the benefits and limitations of intent tagging through a **lab user study** with 12 participants. The study comprised comparative closed-ended and semi-open-ended slide-deck-related tasks that participants completed using our prototype system and, in some comparative tasks, using GenAI features from an existing commercial slide authoring tool. Our findings indicate that users felt more in control and satisfied with intent tag-based interactions than with existing chat-based and design gallery-based generative AI systems for slide deck creation. Participants especially appreciated the **support of non-linear and iterative workflows**, the ability to **express their intent in flexible ways** with **varying levels of ambiguity**, and the **integrated system suggestions** as a valuable and non-distractive aid for **helping them think through the task and figure out what they want** while working on the task.

Building upon these study findings, we illustrate how intent tagging could be generally utilized for facilitating human-GenAI

<i>Intent Elicitation Interface</i>	Open Prompt Format	System Guidance	GUI-based Option Manipulation	Continuous Option Representation
<i>GUI-based / Wizard Dialogues</i>		✓	✓	✓
<i>Text Prompting</i>	✓			
<i>Chat Dialogues</i>		✓		
<i>Intent Tagging</i>	✓	✓	✓	✓

Table 1: Outlining the trade-offs between four intent elicitation methods (GUI-based/Wizard Dialogues, Text Prompting, Chat Dialogues, and Intent Tagging); comparing the factors Open Prompt Format (allowing users to express intent freely), System Guidance (guiding users through predefined options), GUI-based Option Manipulation (adjusting single parameters via GUI elements), and Continuous Option Representation (persistent, single mutable objects).

co-creation interactions across diverse rich content creation tasks beyond slide deck creation. We sketch several **user experience (UX) interface scenarios** for applications such as web blogging, video creation, and 3D scene creation. Based on our prototype system, the study’s findings, and the alternative UX scenarios, we discuss potential design considerations and future work for intent tag-based interactive generative AI systems.

In sum, this paper makes three main contributions:

- (1) *Intent tagging*: graphical micro-prompting interactions to support granular and non-linear co-creation workflows with generative AI systems in the context of slide deck creation and other rich content creation tasks;
- (2) *User study insights on intent tagging for human-AI interaction* in steerability, content creation workflows, and "meta-intent elicitation" by utilizing a GenAI-driven slide creation system called IntentTagger;
- (3) *Design considerations for intent tagging* in GenAI-assisted slide authoring applications in particular, and potentially as a new interaction technique for AI-assisted creative design tasks in general.

Our present work focuses on using intent tags to support the creative design activity of slide deck authoring, which we establish in depth throughout this paper. However, intent tags appear to offer a straightforward design pattern that researchers could explore further in other AI-assisted workflows and scenarios, thereby offering designers a new interaction technique for HCI+AI applications.

2 Related Work

2.1 Intent Elicitation in Human-AI Interaction

In the field of Human-Computer Interaction (HCI), the concept of Human-AI interaction [2] and AI as co-creators [22, 62] highlight both the opportunities and challenges presented by GenAI models. Opportunities lie in automating repetitive and tedious tasks, such as data transformation or formatting, and augmenting creative processes, such as code, text, or image creation [13, 16, 69, 73], while challenges include aligning AI outputs with user intent and users’ uncertainties about effective prompts and AI system’s capabilities [16, 100].

A core challenge in designing effective interactive GenAI-driven systems is ensuring that AI-generated content matches user intentions. This involves two main aspects: the user communicating

their intent to the system and the system interpreting the user’s intent or guiding them in gradually disclosing it (intent elicitation).

Traditionally, **wizard interfaces** have emerged to guide users through predefined option GUI dialogues to elicit their intent step-by-step for complex tasks [20]. However, natural language processing (NLP) advancements have introduced text-prompting-based GenAI models, such as large-language foundation models (LLMs), that allow users to input their intent as **free-form text prompts** [45] or guide users through a **dialogue** to elicit their intents [15, 64, 67, 74].

However, all such approaches present opposing trade-offs (see Table 1): GUI-based menus or wizards can intuitively guide user decisions and reduce ambiguity while allowing for continuous representation of options through graphical manipulatable elements, but they only offer a finite set of options that reflect pre-anticipated use case scenarios by software makers. Free-form text prompting allows users to communicate their intent more flexibly using natural language, but the lack of scaffolding and guidance requires users to know what they want (or need) and how to formulate it as an effective prompt. While easy for simple tasks, communicating intents can be challenging in complex content creation tasks when users might not be aware of all necessary options upfront and are required to iteratively figure these out (intent exploration). As a compromise, natural language chat dialogues can provide system-guided option exploration and scaffolding. However, such systems can only probe on a limited set of options in a reasonable time frame or number of conversational turns, and they often lack continuous option representation and graphical manipulation.

On the other hand, effectively aligning AI-generated content with user intentions also requires awareness of one’s own intentions and understanding of the possible parameter spectrum [89]. Recent work in AI-mediated intent elicitation and sensemaking has explored various strategies and GenAI-driven tools to assist users in more open-ended and iterative ways to elicit and discover their (creative) intent. For example, systems like *Luminate* [85], *Selenite* [44], *Sensecape* [86] and *Graphologue* [36] utilize dynamic LLM-driven GUIs to support users in sensemaking and intent elicitation (e.g., exploring a design space, available options, or topics) by generating and visually clustering semantically related concepts or criteria to help them explore relevant dimensions and parameters matching their tacit intent and needs. Similarly, *SymbolFinder* [65] offers different LLM-driven GUIs to suggest and group related words and images to guide designers in gradually exploring and

finding appropriate symbols to illustrate abstract concepts. Other work by *Kreminski and Chung* [39] uses language models and micro-interactions to elicit users' creative intent by formulating and asking open-ended questions about aspects of a game creator's emerging story world. Similarly, *Germinate* [40], a system for game generation, features an interface to help break down the design intents of casual creators by allowing users to specify game design-related parameters, such as game entities or resources through graphical text-based tags that users can freely define. Based on these user-defined tags, the system attempts to infer additional appropriate tags that users might incorporate or negate (cf. [50]). Similarly, Lin et al. [42] outlined a design space for user interactions with co-creative systems, highlighting the role of explicit communication mechanisms about creative intent, which might be initiated by either the user or the system (mixed-initiative).

From a broader cognitive perspective, self-awareness about one's intentions relates to *metacognition* [27], and recent research by *Tankelevitch et al.* [87] have proposed a stronger focus on supporting users' metacognitive processes, such as *self-awareness* and *task decomposition*, for mitigating challenges of effective GenAI workflows. In a recent design fiction, *Vaithilingam et al.* [90] explore metacognitive support using LLM dialogue-based guided intent exploration inspired by human-human communication patterns such as dynamic grounding, constructive negotiation, and sustainable motivation to support the design process.

Previous work has explored various AI-mediated user intent elicitation approaches and proposed metacognitive user support to improve GenAI-driven tasks. Building atop this prior work, we aim to derive interaction principles for human-GenAI co-creation that combine the benefits of existing intent elicitation interfaces without their drawbacks while supporting users' intent exploration process through metacognitive support.

2.2 Systems for Supporting Prompting and Steering GenAI

Generative AI *foundation* models, such as large language or diffusion models [9, 72], represent a paradigm shift in artificial intelligence, offering task-agnostic pre-training on large-scale data for various downstream applications [77]. Although extremely versatile, research suggests three properties that make interacting with GenAI challenging: *input flexibility* (in handling free-form language, images, code, etc.), *generality* (applicability to a wide range of tasks), and *originality* (ability to generate novel content) [76].

Prior HCI research has documented interaction challenges in prompting and steering GenAI systems across domains such as coding, illustration design, or engineering [33, 43, 46]. For example, users frequently struggle to craft text-based input prompts that will achieve desired outcomes, and face difficulties interpreting and repairing erroneous outputs [100]. Research has begun to explore mechanisms and interfaces to better support users in working with text prompt-based GenAI models. For example, various works have proposed mechanisms to support users in *prompt engineering* [45] through "*prompt augmentation*," which automatically modifies and extends a users' input prompt to improve the model's generated output [7, 11, 81]. While such techniques can improve model output

quality, users are bound to express their intent in an open-ended text format.

Other work has explored ways of *combining direct manipulation interfaces with text-based GenAI systems* to enable more structured user inputs and reduce semantic ambiguity. For example, *Ghost-Writer* [98] offers buttons for predefined prompts for LLM-driven text manipulation tasks. While such approaches reduce misalignment and improve feature discoverability by visually exposing model capabilities, they also restrict the available task options to predefined sets and limit the GenAI models' generality.

Recent work has also proposed contextually bounded prompting mechanisms to offer more fine-grained control over GenAI-driven content generation. *DirectGPT* [54] explores direct manipulation principles for LLMs, such as the continuous representation of manipulatable objects, physical actions to localize the effect of prompts, and reusable prompts in a toolbar. Similarly, *Cococo* [47] offers users a set of GUI widgets for steering GenAI music generation, including slider elements to nudge music generation in high-level directions. Other ideas include painting-like interactions, such as how *Prompt-Paint* [18] enables users to steer diffusion-based text-to-image generation through paint medium-like interactions and *TaleBrush* [19] supports generative story ideation through line sketching interactions to graphically steer LLM-driven story generation.

Building on this work, we contribute to steering interactions for GenAI by investigating mixed interfaces that blend free text input with dynamic graphical elements. We propose intent tagging as graphical micro-prompting interactions to flexibly support intent expression and elicitation using GUI widgets to allow exploration across various levels of ambiguity.

2.3 Non-Linear Content Creation and Iterative Design Workflows

Creating rich content documents, such as blog posts or slide presentations, involves crafting and integrating text, visuals, and multimedia to convey complex ideas while maintaining coherence and audience engagement. Guidelines for creating slide presentations often suggest standardized workflows such as crafting outlines before creating slides [3, 71, 101]. However, other literature emphasizes that slide-creation workflows are less rigid and are mostly shaped by cultural factors and organizational norms [97]. For example, in organizations, people often start or continue presentations from different starting points, such as from existing documents, other slide decks, or templates.

Rich content creation processes are also iterative and diverse in nature, typically progressing through cyclically occurring stages [5, 23], where creators alternate between '*exploration*' and '*exploitation*' in a sense-making process to reach a final outcome [66]. Content creators continuously develop, reflect on, and act on their ideas and plans [38, 78]. The interaction between users' perception of the material and the material itself enables creators to develop, reflect on, and question their understanding, leading to new ideas and improved plans [37]. Research has underscored the role of reflection in content creation as a necessary moment for creators to situate their ideas and plans in the appropriate context [24, 61, 80].

However, existing challenges around prompt formulation engage users in cognitively demanding trial-and-error processes and hinder

reflection related to content creation. Furthermore, many of today's GenAI systems assume a specific workflow or require users to adapt their processes to integrate GenAI effectively [75, 87].

In conclusion, despite recommended workflows for content creation (such as starting with an outline), content creation processes are typically highly situational, non-linear, and iterative in nature. Therefore, we aim to derive interfaces for GenAI-driven content creation systems to promote non-linear and diverse workflows while supporting reflection on the content creation process.

2.4 Systems for Supporting Slide Deck Creation

Creating slide presentations is a multifaceted task that involves various sub-tasks, and previous research has introduced a range of approaches to support different stages of this process.

Besides popular slide-by-slide authoring tools like PowerPoint or Keynote, research has suggested interfaces to support users in **iterative slide creation workflows** by managing multiple presentation versions [25], prototyping slide decks using markup language [26] or authoring slide presentations on a 2D canvas interface [41] for supporting non-sequential workflows.

Besides manual authoring systems, numerous works have explored ways to **automatically generate slide presentations** and rich content documents for single topics [95], or scientific [30], technical [48], educational [21], or semantically annotated source documents [52]. Other systems focus on automatically generating presentation-specific aspects, such as visual-textual layouts [96]. While these systems allow for automatic content transformation and slide creation, they lack methods for users to influence or control the generation process.

Several works have, therefore, proposed mechanisms to enable users to **generate presentations in more controllable and workflow-integrated ways**, such as allowing data scientists to steer the generation of slide presentations from Jupyter notebooks by interactively selecting code cells and slide layouts [93] or through a user-controllable text outline [92]. Similarly, *Knowledge-Decks* [17] allows users to generate slides documenting data science knowledge-discovery processes from automatically collected user behavior events from visual analytic tools.

Previous work has explored various authoring interfaces, automatic slide generation from source documents, and user-controllable slide generation in domain-specific applications like data science. Expanding on these approaches, we aim to explore interaction principles for GenAI-supported slide deck creation across diverse content tasks and workflows. Our approach combines GenAI-driven slide generation from multiple input types (e.g., prompts, text documents, images) with *non-predefined* and *non-linear authoring workflows* in mind that let users seamlessly switch between generation and editing modes across outline, individual slide, or entire deck level.

3 Exploring Design Principles for GenAI-supported Rich Content Creation

In this section, we lay out the challenges of rich content document creation in the context of slide creation and how these informed the design principles that led us to intent tagging as an interaction paradigm for enabling novel human-GenAI co-creation workflows.

Slide deck creation is a complex task that requires balancing multiple facets, such as narrative coherence, visual style consistency, and the integration of diverse content sources [3, 71, 97]. These elements must not only align individually but also work together harmoniously to deliver an impactful presentation. The challenges arise from the need to manage and sequence multimedia content while simultaneously crafting a narrative that flows logically across slides. Achieving this requires careful consideration of storytelling, visual design, and the inclusion of relevant data or references, often making the task overwhelming, especially under time constraints or when dealing with open-ended creative briefs.

One of the core difficulties in slide deck creation is the need to manage these distinct yet interdependent elements—*narrative*, *visual style*, and *content sources*—without losing sight of the overall presentation goals.

Traditional tools often compartmentalize these tasks, forcing users to switch between different modes or tabs, which can disrupt the creative process. Linear workflows or predefined templates, while useful in some cases, often fail to accommodate the nuanced adjustments needed to refine a slide deck to a professional standard.

Moreover, as discussed in the previous section, the integration of Generative AI (GenAI) into slide deck creation introduces both opportunities and challenges. Current GenAI approaches tend to offer either (semi-)automated solutions with little ability for users to steer the content generation, design galleries with limited predefined option sets, or chat-based interactions that can feel too linear, unstructured, and limited for exploring alternatives.

3.1 Design Principles

Based on these requirements, we developed the following interaction design principles for GenAI-supported slide creation:

DP1 Enabling flexible, non-linear, and iterative workflows

To cater to users' diverse slide creation needs and individual working styles, users should be empowered to start and refine their presentations in a variety of ways. Whether beginning with an outline, building on an existing deck, crafting a single slide, or incorporating content from other documents, users should have the flexibility to approach their tasks from any angle. Additionally, the interface should allow seamless transitions between different content views—such as slide, deck, or outline—enabling GenAI-assisted iterative enhancements at any stage of the creation process.

DP2 Accommodating diverse steering input types

Slide creation often requires the integration of multiple sources of information, ranging from explicit, well-defined data (e.g., an annual sales report) to more abstract, nuanced considerations (e.g., audience-specific messaging). To ensure that GenAI-generated outputs align closely with user intentions, the interface should facilitate the input of a wide range of content types, including natural language descriptions, existing documents, images, and other relevant media. By enabling users to provide rich, varied inputs, the system can better interpret and fulfill the specific needs of each presentation.

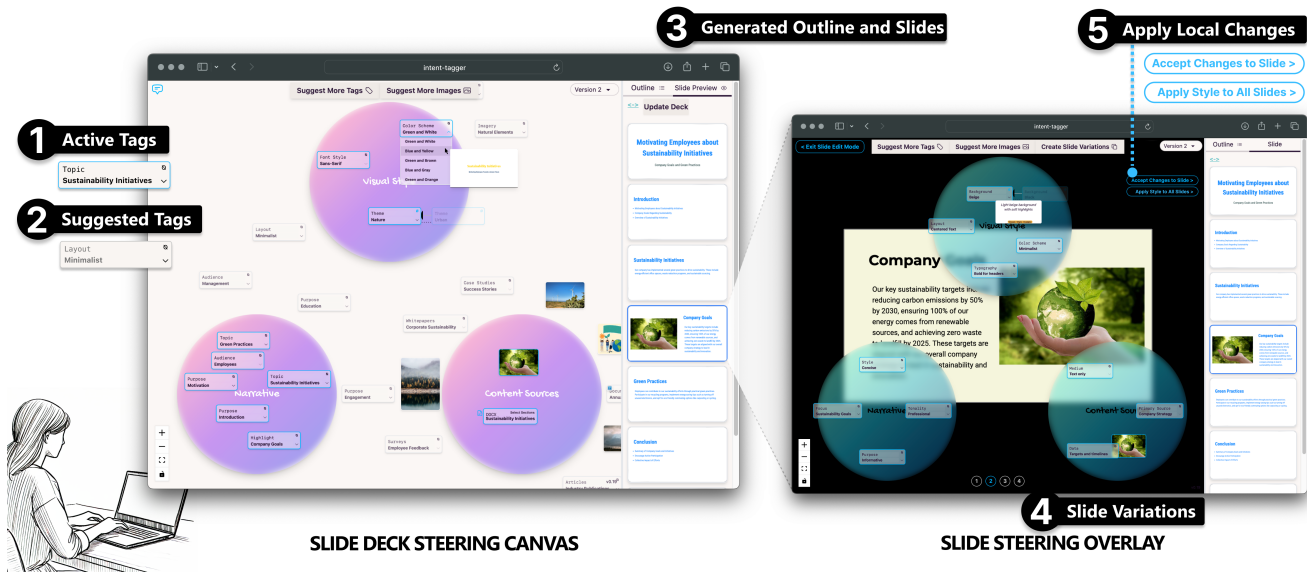


Figure 2: Screenshots of the *IntentTagger* system: (Left window) A user creates a slide presentation using the Slide Deck Steering Canvas with (1) active tags in the three circular tag groups surrounded by (2) inactive system-suggested tags. (3) The generated outline and slides are displayed in the tabbed sidebar. (Right window) The user makes (scoped) adjustments to a single slide via the Slide Steering Overlay and (4) explores generated slide variations. (5) Then, local slide changes can be accepted or applied to the entire deck.

DP3 Supporting intent expression across varying levels of abstraction

In many slide creation scenarios, users may have clear, specific ideas about certain aspects of their presentation, while other elements remain less defined or harder to articulate. To accommodate this variability, a GenAI interface should allow users to express their intent at different levels of abstraction—ranging from detailed, precise instructions to more general, high-level directives. This flexibility not only helps users articulate their ideas more effectively but also supports iterative exploration and refinement of their presentation content.

DP4 Leveraging AI for contextual content and terminology suggestions

To assist users in overcoming challenges like the "blank page syndrome" or finding terminology to describe their intent, the interface should harness the associative power of LLMs to provide contextual suggestions and inspiration. By offering alternative ideas, vocabulary, and creative prompts, the system can stimulate users' thinking, helping them to refine their content and explore new directions in an iterative process.

DP5 Providing pre-computed real-time previews

To allow users to better and quicker anticipate how specific choices will impact generated slides, the tool should provide pre-computed real-time previews to, for example, let users rapidly explore how different fonts or colors would look on a given slide.

4 IntentTagger: An Intent Tagging-based Slide Creation System

Driven by the requirements of rich content creation tasks and the design principles described in the previous section, we explored novel input strategies through graphical micro-prompting interactions—an interaction notion we coined *intent tagging*. In addition to our design principles, a central inspiration for conceptualizing creative intent elicitation through granular and flexible micro prompts that represent users' intention was the "mood board" technique, in which creatives iteratively compile visual and textual elements into a collage to convey an overarching theme or creative direction [31]. Similar to how mood boards enable capturing and expressing designers' intent through compiling different media with varying levels of abstractions—such as words and pictures ranging from abstract concepts, stylistic references, or specific material details—intent tagging allows users to express their intent openly through collections of intent tags.

To explore further the idea of intent tagging, we created *IntentTagger*, an LLM-driven system allowing users to iteratively create and modify slide deck presentations using intent tags on a 2D canvas interface (see Figure 2). In this section, we demonstrate the utility of *IntentTagger* by first illustrating its functionalities in a use-case scenario. We then follow with a detailed description of the proposed interface mechanisms and their implementation.

4.1 Example Use Case Scenario

Lucy is a sustainability manager at a company and wants to prepare a short slide presentation for new employees to inform them about sustainability initiatives at her company.

Lucy starts IntentTagger and clicks on the "Create a new presentation from prompt..." button. A modal popup appears, and she types "a presentation for introducing sustainability initiatives for new employees" in the prompt field and presses enter. The modal disappears, and the three circular steering groups on the steering canvas pulse, then three concept tags appear in the "Narrative" circle with the parameters "Topic: Sustainability Initiatives," "Audience: New Employees" and "Purpose: Introduction."

Tag Suggestions: To get inspiration for further presentation talking points, she clicks "Suggest More Tags," and shortly after, new tags appear outside of the circular tag groups. Lucy finds some of the suggested tags useful and drags these into the "Narrative" group to include them in the generation, such as "Focus: Eco-friendly Practices," "Highlight: Company Goals," "Section: Key Initiatives," and "Objective: Corporate Responsibility."

Including External Content: Lucy also wants to include sections from an existing Word document detailing the company's policies. With the mouse cursor, she drags the document from her file browser onto the IntentTagger app, and a new reference tag appears in the "Content Sources" group. She clicks on the tag's "Select Sections" button, which opens up the tag's Text Selection Widget showing the document's content. In the widget, she highlights two document sections, clicks "Include Section in Presentation" from the widget's context menu, and closes the widget.

Generating an Outline: Next, she wants to generate a first draft of the presentation's outline and presses "update outline from intent tags" in the outline view panel. The tag groups start pulsing, and then the new outline appears in the outline text editor.

Manually Crafting Concept Tags: Lucy is happy with the structure overall but finds it too long (20 sections). Instead of manually editing the outline, she adds a new concept tag to the "Narrative" group and enters "Number of slides" into the tag's upper text field and "10" into the lower and clicks the "update outline from intent tags" button again. The newly generated outline is now shorter and only comprises ten sections.

Generating Slides: Next, Lucy wants to generate a first draft of the slide deck and clicks the "Update slides from intent tags" button. Shortly after, the generated slides appear in the slide preview side panel. After evaluating the slides, Lucy realizes that adding some images and adjusting the visual style could enhance the presentation's appeal.

Including Suggested Images: She clicks the "suggest more images" button on the tag steering board, and shortly after, new Reference Image Tags appear outside the "Content Sources" tag group with eco-friendly-themed stock photos. Lucy drags some of the image tags into the "Content Sources" group to add them to the presentation.

Adjusting the Visual Style through Tags' Alternatives Drop-down: To make visual adjustments to the slides, she drags in some of the previously suggested tags, such as "Typography: Sans-Serif," "ColorPalette: Green and Blue," and "Theme: Nature" onto the "Visual Style" tag group. To get a better sense of typography alternatives, she clicks on the tags' drop-down, and a list of alternative tag values appears, such as "Serif," "Monospace," and "Handwritten." While hovering over each option with the mouse, Lucy sees a preview thumbnail of that font style applied to the current slide. Lucy chooses "Monospace." Now, Lucy clicks "Update slides from intent

tags" again, and shortly after, she sees the revised slides reflecting the intended new visual style, including the images.

Adjusting a Single Slide with Tag Steering Overlay: Lucy decides to refine the slide deck further. She clicks on the third slide in the Slide Preview panel to open the slide edit view and review its content in detail. The slide contains a text paragraph and an image, but she wants to show a list of bullet points instead. To modify the slide, she clicks on the intent tag icon next to the slide, and the three steering tag groups appear as an overlay containing tags representing the slide's Narrative, Visual Style, and Content Sources. Lucy selects the "Text Format" tag inside the Narrative group and clicks on its drop-down widget. A list of alternative tag values appears, including "Bullet Points," "List," and "Table." Lucy chooses "Bullet Points." To adjust the background color, Lucy creates a new tag in the Visual Style group and types in "Background" and "pastel color." Now, she clicks on the "Explore slide variations" button, and the system generates several alternative versions of the slide, each with slightly different text variations and pastel background colors. Lucy clicks through these variations and selects one that better aligns with her presentation's tone.

Applying a Slide's Style to All Slides: To maintain consistency across the entire presentation, Lucy uses the same visual style and selects "Apply to all slides."

Satisfied with the changes, Lucy saves the presentation, ready to deliver it in the upcoming meeting with her colleagues.

Adjusting the Presentation to New Requirements: After a couple of months, Lucy needs to adapt the presentation for a different audience—existing employees rather than new hires. She clicks on the "Audience" tag in the Narrative group and changes it from "New Employees" to "Existing Employees." The system instantly updates the content and tone of the slides to better fit this new audience. For example, introductory sections are replaced with more in-depth discussions of ongoing sustainability projects and their impact.

4.2 Interface Features

Here, we provide a more detailed explanation of IntentTagger's features:

4.2.1 Intent Tags: Concept Tags, Reference Tags and Groups (DP2). Intent tags are collections of single keywords, phrases, or media artifacts describing users' intended outcomes. Tags serve as granular and flexible micro prompts on a zoomable 2D canvas. Users can steer the generative slide creation system through two types of tags:

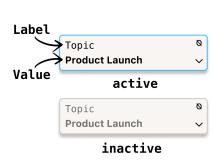


Figure 3

(1) **Concept Tags:** Concept Tags are two-part text micro prompts of the format [label: value], such as "Topic: Product launch" or "Color scheme: Corporate." Tags can be active or inactive to toggle their influence on generated content.

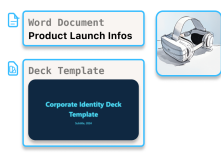


Figure 4

(2) Reference Tags: *Reference Tags* allow users to provide external documents such as Word documents, images, or other slide decks as references for slide generation. Users can add these files by drag-and-drop from their file browser.

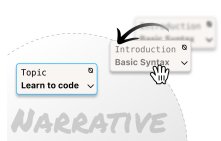


Figure 5

Tag Groups: All active tags are visually clustered into three circular groups on the interface, each representing different slide creation aspects: *Narrative*, *Visual Style*, and *Content Sources*. Users can drag tags in and out of groups to activate and deactivate them.

4.2.2 Deck Steering Canvas and Slide Steering Overlay (DP1, DP3). Users can generate and modify single slides or entire slide decks using Intent Tags. The *Deck Steering Canvas* allows users to steer the (re)generation of entire slide decks. For example, specifying the number of total slides or changing the overall tone of the deck's text from brief to more verbose. Alternatively, the *Slide Steering Overlay* allows users to steer the (re)generation and explore alternatives of single slides.

4.2.3 Outline Editor (DP1). Parallel to slides, the system also generates a presentation outline, which can be optionally edited using the outline editor. This editor also allows participants to start a new presentation by first drafting an outline or letting the system generate an outline from provided intent tags.

4.2.4 Editing Intent Tags, Tag Widgets, and System Suggestions (DP3, DP4, DP5). Intent tags allow users to formulate their intent in several ways, such as manually creating and editing tags or choosing from options generated through LLM-driven dynamic adaptive UI mechanisms:

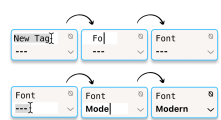


Figure 6

Manually Crafting Tags Users can manually create and modify tags by freely typing words or phrases into its two text fields, such as "Font: Modern" or "Tonality: Engaging."

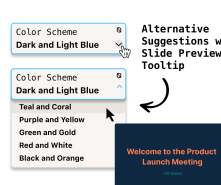


Figure 7

Drop-Down List Users can also explore alternative properties for each concept tag by browsing its dynamically generated drop-down list, which also provides pre-generated visual previews in a hover tooltip.

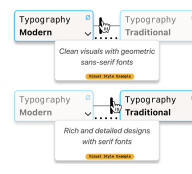


Figure 8

Opposite Slider Widget Users can fine-tune tag values with more granularity by using a slider widget. For this, the system automatically generates an opposite value for each tag with a five-step continuous range slider in between, including descriptive examples illustrating each step's value.

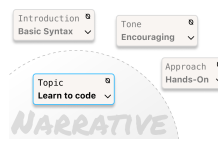


Figure 9

Tag Suggestions For each tag group, users can request new tag suggestions, such as additional talking points, visual styles, or content sources, from the system based on already specified tags. Generally, all suggested tags first appear outside the circular groups, and users can drag suggested tags into a group to include them in the generation.



Figure 10

Image Suggestions In addition to tags, users can also request system suggestions for new image tags. When requested, each time the system returns different images from a background online image search related to the board's active tags.

4.2.5 Steering Slide Generation. To summarize the different ways of steering slide generation with Intent Tags, Figure 11 illustrates how including a tag suggestion modifies the content of a slide, using the drop-down list changes a slide's mood, and using the opposite slider widget blends between a text- and image-heavy layout.

4.2.6 Previews of Slides and Slider Values (DP5). IntentTagger integrates several preview features, allowing users to explore alternative tag values without committing to long waiting times: The tag's *Drop-down* list shows pre-generated previews as tooltips of how the current slide would look like with that tag value applied (Figure 7). The *Opposite Slider* widget shows generated explanations describing the resulting effect if applied to the generation (Figure 8). All tag previews and explanations get asynchronously pre-generated in the background, allowing users to explore these options in real time.

4.2.7 Referencing External Documents (DP2). Besides *Concept Tags*, users can also include external files for the generative slide creation as *Reference Tags*.

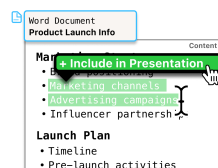


Figure 12

Including Content From Word Documents Users may drag in Word documents as content sources for the generated presentation. Optionally, users can select sections through a text selection pop-up to include only specific parts in the generated presentation.



Figure 11: Examples of steering GenAI slide generation with Intent Tags: (A) Influencing the detail level of the displayed information, (B) steering the "mood" from "bright" to "dark," or (C) gradually blending the layout between text- and image-heavy.

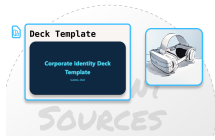


Figure 13

4.2.8 Tag Grounding Acts (DP4). While Intent Tagging presents granular and flexible mechanisms for users to communicate their intentions to the generative AI slide creation system, this principle also works backward in so-called *Tag Grounding Acts*, where the system creates intent tags from user inputs such as longer text prompts or slides.

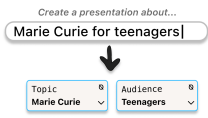


Figure 14

Grounding from Text Prompt This feature allows users to start a new presentation by providing instructions via a longer conventional text prompt, which the system will then automatically decompose into individual Intent Tags.

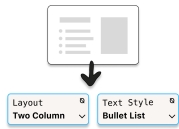


Figure 15

Grounding from Slide Each time a user invokes the *Slide Steering Overlay* to adjust single slides, the system first analyzes the slide's *content sources*, *narrative*, and *visual style* and then pre-populates the interface with intent tags representing the slide's current state. This provides an easy starting point for users to make adjustments through Intent Tags.

4.3 Implementation Details

IntentTagger is implemented in TypeScript using ReactJS [70] with ReactFlow [28] for the tag canvas interface and Blocknote [8] for the text editor features. Mommth.js [94] is used for importing Word documents, and slide rendering is based on a modified fork of spectacle.js [29]. For generating slides and adaptive UI, we use the official OpenAI API [63] to execute prompts using "gpt-4o" and the Bing search API [56] for image suggestions. To suggest new tags, we prompt GPT to generate a list of *[label:value]* pairs for each tag group to augment the existing set of active intent tags. For outlines, GPT is instructed to generate a presentation outline in markdown

format from a list of a user's active concept and reference tags (including images if present). To generate slides, we then prompt GPT to return a template-based JSON slide deck representation by providing the outline, active intent tags, and deck references (if specified). For generating the drop-down values, slide previews, and slider explanations, the system asynchronously requests these from GPT in the background for each new or modified tag on the canvas. Please see the Appendix for detailed prompts.

5 User Study

To better understand the possible benefits and limitations of intent tagging-based interactions, we conducted a lab user study aimed at providing insights into these research questions:

- RQ1** *What are the key differences between chat-based and intent tag-based interactions with GenAI?*
- RQ2** *How do people work with intent tags?*
- RQ3** *What are users' perceived benefits and challenges for GenAI-driven slide creation with intent tags?*

For RQ1, we decided to compare intent tag-based interactions with chat-based and design gallery-based approaches (text prompting and choosing from a set of options) since these represent the currently most common forms of interacting with GenAI for content creation in commercial systems, such as in OpenAI's ChatGPT/DALL-E [7], Adobe's Firefly [1] or Microsoft's PowerPoint Copilot [58] and Designer [59] features.

5.1 Participants

We recruited 12 participants (*8 self-identified as females, 3 males, and 1 non-binary, age $M=31.6$ years ($SD = 7.65$)*) with professional slide presentation creation experience via email lists at a large software company. The participant pool consisted of individuals with diverse job titles, and the majority of selected participants used PowerPoint at least multiple times per month in their jobs (see Table 3 in the Appendix). All participants had previous experience using the *Designer* and *Copilot* features in PowerPoint, ensuring existing familiarity with generative AI functionalities for slide deck creation. Participants signed an IRB-approved consent form and were compensated with a \$50 gift card after study completion.

5.2 Procedure and Tasks

The lab study was structured into four phases:

- 1) On-Boarding (20 min):** At the beginning of the session, after a general study introduction, participants watched a video tutorial demonstrating IntentTagger's core functionalities with a

step-by-step example. Following the video, participants were asked to complete two five-minute guided hands-on structured tasks (editing an existing slide and making adjustments to an entire deck) to familiarize themselves with the tools' interface and operation.

2) Comparative Tasks (2 x 10 min): After the onboarding phase, participants completed two comparative tasks, each lasting 10 minutes, to evaluate their ability to create presentations using IntentTagger and a baseline system. In both tasks, participants were asked to create a 6-slide presentation aimed at educating teenagers about the inventions and scientific discoveries of a historical figure (*"The Discoveries of Marie Curie"* or *"The Inventions of Nikola Tesla"*). Participants alternated between using IntentTagger and Microsoft PowerPoint [57] across the two tasks in randomized order. For PowerPoint, participants were restricted to using only the integrated Copilot feature [58] (chatbot interface with optional document upload) and the Designer feature [59] (design gallery for slide layouts) for slide generation and modification. As a starting point, participants were provided with a Word document containing the relevant Wikipedia article in both tasks. Participants were required to think aloud during the tasks, and after completing each task, they filled out a survey with attitudinal 6-point Likert scale questions. To mitigate order effects, the order of systems and presentation topics was randomized across participants.

3) Semi-structured Task (10 min): In the third phase, participants were tasked with creating a 7-slide presentation from scratch using IntentTagger. They chose a topic related to a hobby they enjoy, aiming to convince others of its value and explain how to get started. Participants had ten minutes to complete the task, using only the prototype system without directly editing text or images on the canvas. They were required to think aloud as they worked, and before starting, they briefly described their topic and intentions for the presentation. We created this task to encourage participants to engage more freely with IntentTagger's features, focusing on content creation and presentation design related to a topic they are knowledgeable about and emotionally connected to. Since IntentTagger's deck generation time increases per slide, we deliberately limited the number of slides to six and seven per task in phases 2 and 3 to keep the processing time for each cycle within 15 seconds.

4) Exit Interview (20 min): In the final phase, participants participated in a semi-structured interview to provide feedback on their experience with IntentTagger, focusing on its overall utility, comparison with tools like PowerPoint's Copilot, and the effectiveness of intent tagging interactions for slide creation. They shared insights on the tool's strengths, areas for improvement, and potential integration into their professional workflows.

5.3 Collected Data, Measures, and Analysis

Across the study, we collected the following data:

- Video, screen, and audio recordings and machine-generated transcripts of the task think-aloud sessions
- Audio recordings and machine-generated transcripts of the post-task interviews
- Post-task survey data
- Participant-created presentations and related IntentTagger project files

To compare chat-based and intent tag-based interactions (Q1), we analyzed the post-task surveys from phase 2 that probed on participants' perceived ease of use, efficiency, and control over the slide generation process on a 6-point Likert scale. We applied the Wilcoxon signed-rank test to assess statistical significance and calculated 95% confidence intervals for mean differences via bootstrapping with 10,000 replications using R [68]. This approach has been suggested for similar data and studies [53, 102].

To answer how people work with intent tags (Q2), we conducted a video interaction analysis [6] of the video recordings collected in the semi-structured task (*study phase 3*). We manually coded participants' interactions with IntentTagger's features in Atlas.ti [4], such as their interactions with tags or the moments they triggered the outline and slide generation.

Finally, to investigate users' perceived benefits and challenges of intent tags (Q3), we conducted a reflexive thematic analysis [12] of the interview transcripts. We followed an iterative inductive coding process (using Marvin [51]) and generated themes through affinity diagramming using Miro [60].

6 Study Findings

Participants went through the tasks without abandoning them.

6.1 Key differences: Chat vs Intent Tag-based Interactions (RQ1)

6.1.1 Observations. In the two comparative tasks (phase 2), participants had to create a presentation about Nikola Tesla or Marie Curie on six slides using engaging images and language suitable for teenagers, based on a provided Word document containing information from a Wikipedia article. Here, we report on the observed key differences between participants' processes during these tasks when working with the chat-based system and IntentTagger.

Chat-based System. In the chat-based system, participants typically began by providing a straightforward task prompt, such as *"Create a new presentation from /Tesla.docx about Nikola Tesla for teenagers on six slides."* However, the system consistently generated a 20-slide PowerPoint presentation based on all content from the document in a generic way, including generic and vaguely related stock images. When participants tried to refine their prompts, such as specifying the desired number of slides or requesting specific content to be highlighted, the system frequently ignored these requests and generated a new slide deck similar to the previous one. As a consequence, participants found themselves repeating their instructions or trying different phrasings to guide the system, often without noticeable changes. The system frequently returned messages like, *"I don't recognize that wording as something I can do,"* offering alternative actions that were not relevant to the task. Consequently, participants spend most of their time trying out different text prompts to get the system to generate a more targeted, shorter presentation; however, in all cases, without success. Occasionally, participants went on to tweak the layout and visual style of single slides using the Designer feature by selecting a layout from a design gallery list. While they managed to adjust

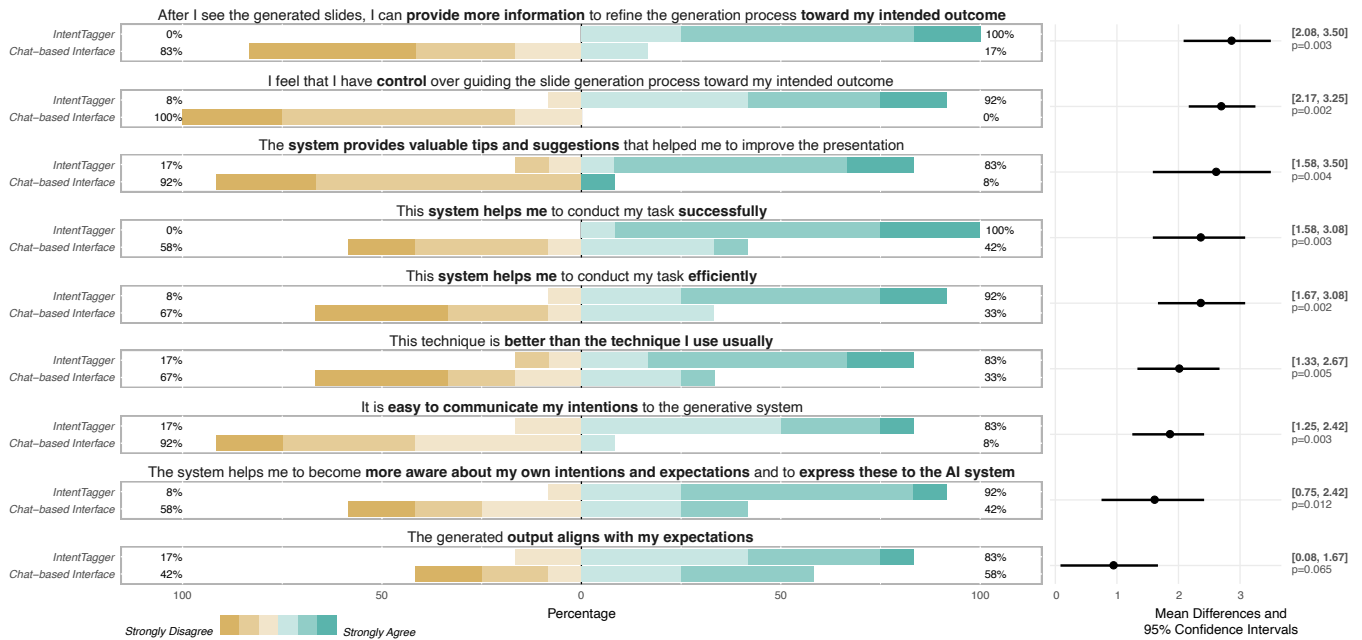


Figure 16: (Left) Participants’ responses when rating the 6-point Likert statements for IntentTagger and Chat-based Interface, ranked from largest to smallest effects; (Right) Dots show the mean difference of IntentTagger compared to Chat-based Interface; Bars are the 95% CIs calculated with the studentized bootstrap method.

the style of single slides in isolation, they could not achieve a consistent look and feel for the entire deck this way.

IntentTagger With IntentTagger, participants typically began by providing the Word document as a reference tag and then continued by providing further concept tags for key elements like the number of slides, audience, and the focus areas of the content. In addition, many participants used the reference tag’s text selection widget to specify specific sections of the provided Word document to ensure that only relevant information was included in the presentation. After specifying tags matching the task brief, most participants also made use of the image suggestion feature from which they added some images to the content source tag group. Then, after generating slides, they either continued to make refinements on single slides and then updated the deck’s style based on the adjusted slide, or they continued refining the entire deck by adding or modifying intent tags on the deck steering canvas. Participants also frequently used the system’s tag suggestions to refine the presentation’s content and visual style.

From a usability perspective, some participants struggled initially with understanding how to effectively use the intent tags, particularly in determining how granular or broad their tags should be or to which tag group certain tags would belong. However, they soon became more confident and many participants commented on the ability to better convey their intent with intent tags and tag groups: “it definitely feels like I get something close to what I had in mind [...] Looks like something efficient, I will be happy to use it in my work.” (P09)

6.1.2 Questionnaire. In the comparative post-task surveys, participants rated all nine 6-point statements higher for Intent Tagger (Figure 16). Participants expressed a **significant improvement in intent elicitation and alignment with IntentTagger** over the chat-based system regarding the **iterative process** (“After I see the generated slides, I can provide more information to refine the generation toward my intended outcome”, MD = 2.83), **control over guiding the system** (“I feel that I have control over guiding the slide generation process toward my intended outcome”, MD = 2.67), **intent communication** (“It is easy to communicate my intentions to the generative system”, MD = 1.83), **meta intent elicitation** (“The system helps me to become more aware about my own intentions and expectations during the task and to express these to the AI system”, MD = 1.58) and **alignment with their expectations** (MD = 0.92). Participants also valued the **guidance for slide creation** of IntentTagger higher (“The system provides valuable tips and suggestions that helped me to improve the presentation”, MD = 2.58) and felt they were able to **conduct the task more successfully** (MD = 2.33) and **efficiently** (MD = 2.33) than with the chat-based system. They also felt that IntentTagger would be **better than the technique they usually use** (MD = 2.0).

6.2 Observed Content Creation Workflows Using Intent Tags (RQ2)

During the semi-structured task (study phase 3) participants created a presentation from scratch about a personal hobby by using IntentTagger within 7 – 12 minutes (see Table 2). Most participants started by using the *Grounding from text prompt* feature to formulate an

Table 2: Overview of semi-structured task results (phase 3); reporting the self-chosen topic of the slide deck and task process summary statistics; *Note: During P04’s session, the image suggestions failed to load due to an API outage.

ID	Slide Deck Topic	Task Time (mm:ss)	# Outlines Generated	# Decks Generated	# Total Tags	# Concept Tag Sugg. Included	# Img Tag Sugg. Included	% Sugg. Included	# Drop-Dwn Included
P01	Hiking	11:19	3	4	14	3	4	50%	3
P02	Yoga	06:32	3	3	21	12	5	81%	2
P03	Kayaking	07:01	3	2	25	13	5	72%	1
P04*	Homemade Pizza	08:55	2	3	20	17	0*	85%	4
P05	Growing dahlias	08:12	2	5	16	6	5	69%	1
P06	Racing road bikes	08:20	2	1	14	4	3	50%	2
P07	Skateboarding	07:53	1	3	12	8	3	92%	2
P08	Dungeons and dragons	06:42	2	1	21	9	6	71%	0
P09	Practicing Taekwondo	11:43	5	3	11	5	3	73%	2
P10	Tennis	09:01	3	4	16	8	3	69%	0
P11	Why trombone is cool	09:07	3	2	16	8	4	75%	0
P12	Paddle boarding	09:07	2	2	24	17	6	96%	2
Mean		08:39	2.6	2.8	17.5	9.2	3.9	73.5%	1.6

initial longer text prompt, which the system decomposed into an initial set of intent tags from where the participants continued to work on the Deck Steering Canvas (see Figure 17 upper lanes). Only two participants (P03 and P10) decided to start by adding concept tags manually. On average, at the end of their sessions, users had 17.5 active intent tags (concept and image tags combined) and had generated 2.6 outlines and 2.8 slide decks. Of those tags, on average, 9.5 were tags users took from the system’s suggested concept tags, and 3.9 were image tags (ranging between 50–96% of active tags that users chose to include from system suggestions). Optionally, we included screenshots in the Appendix to give an impression of the variety of participant-created slide decks and tag steering boards.

6.2.1 Workflows. All participants generated an outline before generating a slide deck (see Figure 17 upper lanes). While this order was partially driven by IntentTagger’s requirement for an outline to enable slide deck generation, many participants also expressed in the post-task interviews that this order felt intuitive to them (see section 6.3.2). Almost all users first specified several intent tags for the narrative before generating a first outline after 2–4 minutes. However, two users (P05 and P08) generated a first outline before making substantial tag edits and then refined these using tags. Most participants did exclusively steer the outline and slide deck generation using intent tags, while three users (P04, P07, P08) made manual edits to the outline at a point. In addition, most users worked exclusively in the Deck Steering Canvas, but three users (P09, P11, P12) also used the Slide Steering Overlay to adjust single slides and then updated the deck’s style from that slide. Overall, a common pattern was adding or editing intent tags in the Narrative tag group and then optionally generating a first outline and slide deck. From there, most people continued adjusting the visual style through tags, and lastly, they added image tags to the content sources group.

6.2.2 Prompting With Intent Tags. Participants demonstrated a variety of approaches when prompting with intent tags, particularly when dealing with ambiguous or open-ended tags like color or style. For example, one participant specified the header color as “Gold” and the color scheme as “Dark and Moody” (P11), while another participant used hexadecimal notation (“#FF33CC”) for body color (P05).

Users’ Initial Uncertainty Around Tag Naming

Despite the flexibility, some participants experienced uncertainty about how to “correctly” name certain tags, particularly when it came to specifying the number of slides. The task required a seven-slide presentation, and users created tags like “Number of Slides: 7,” “Slide number: 7 slides,” and “length: 7 slides in total.” During the think-aloud sessions, participants expressed hesitation about whether they were using the right terminology. However, after seeing the generated slide deck, they gained confidence in the system’s ability to interpret a variety of tag expressions effectively.

Users’ Initial Uncertainty Around Tag Placing

Interestingly, while most participants placed the tag for the number of slides within the narrative tag group, there was some variation in how users organized their tags. For instance, one participant placed the slide number tag under visual style (P08), while another placed it in the content sources group (P10). Regardless of where the tag was placed, the system consistently produced a seven-slide presentation. This adaptability underscored the system’s capacity to handle user input flexibly, ensuring that the end result matched the users’ intent even when the tags were categorized differently.

Flexible Handling of Longer Prompts

One participant created an exceptionally long and detailed prompt when using the initial grounding through text prompt, which led to the generation of tags containing chains of comma-separated phrases. While these tags were visually less ideal and

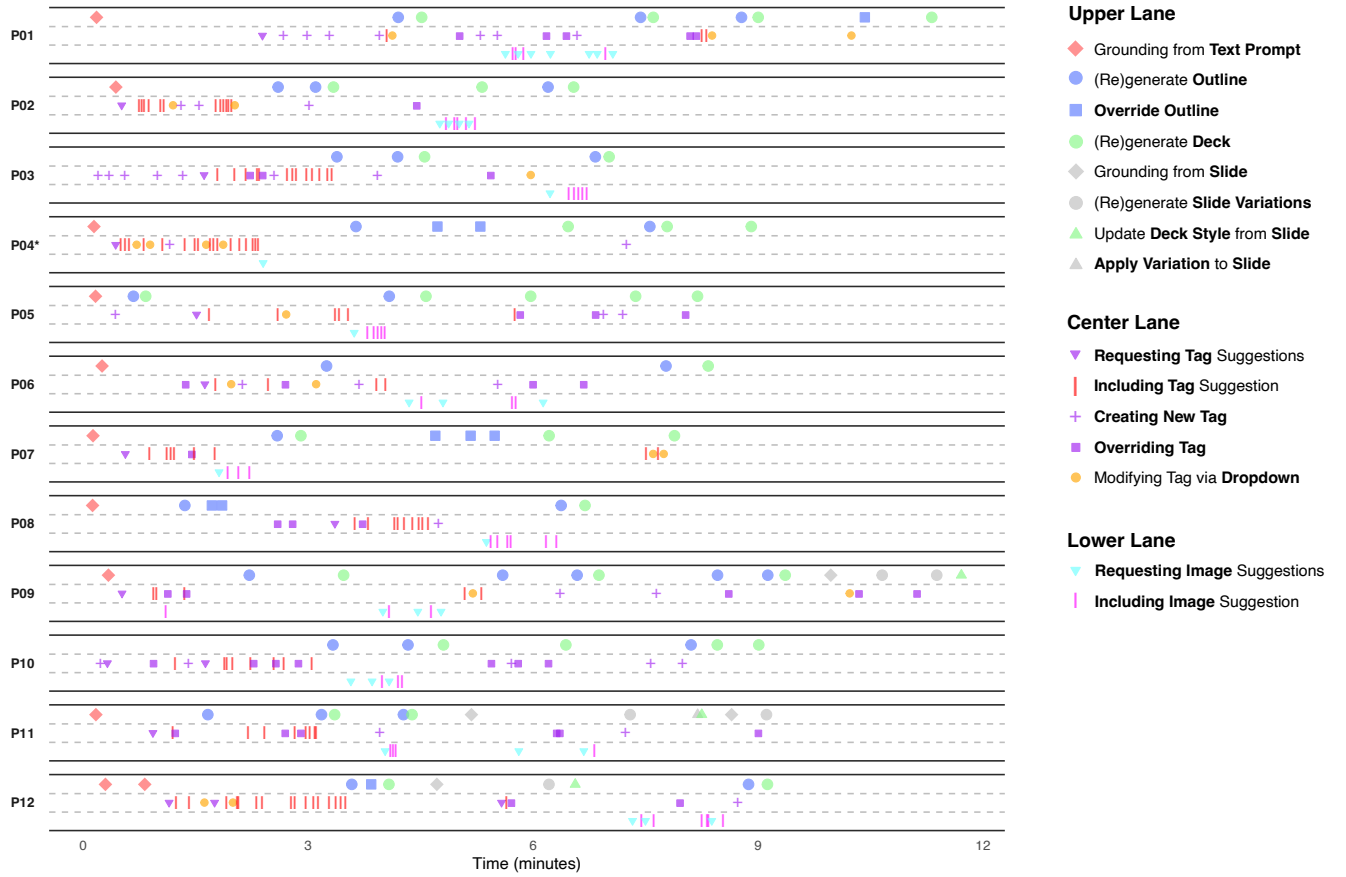


Figure 17: Timeline plots visualizing participants’ interaction events of the semi-structured task (phase 3); upper lanes show events related to outline and slide generation; center lanes show interactions with concept tags; lower lanes show interactions with image reference tags; tag modifications via the *Opposite Slider Widget* are omitted since only a few participants occasionally explored slider options but without applying these to the slide generation (see section 6.2.3) *Note: During P04’s session, the image suggestions failed due to an API outage.

cluttered the workspace, the system still managed to process them effectively (P08).

User-Driven Tag Hierarchy and Organizational Strategies

In addition to varying naming conventions, some participants developed a sense of structure and hierarchy by visually organizing the tags within the tag groups. Although the generative system currently does not inherently recognize such tag hierarchies, it was interesting to observe how participants imposed their own sense of order. Another strategy for creating hierarchy through tags was observed when P01 explicitly labeled tags to represent an ordered flow, such as “*Topic 1: hiking*” and “*Topic 2: hiking shoes*.”

Overall, the variety of tag naming and placement strategies showed that while some users were initially unsure about the “right” way to prompt the system, they quickly became more comfortable with the process and how to utilize the tag’s granularity and various levels of ambiguity.

6.2.3 Interactions With System Suggested Tags. Participants actively engaged with system-suggested tags early in their content

creation process, with all users requesting tag suggestions within the first three minutes. This was typically followed by an intense phase of tag selection and editing that lasted 1 – 3 minutes (see clusters of red vertical strokes ||| in Figure 17). For example, in a presentation about road cycling, the system suggested visual style tags like “*fonts: sporty and dynamic*” and “*backgrounds: scenic roadways*,” which the participant included (P06). Similarly, in a kayaking presentation, the system suggested “*Font: Casual and Readable*” and “*Icons: Outdoor activity symbols*,” demonstrating the system’s ability to offer contextually relevant suggestions that aligned with participants’ content themes (P03).

Users leveraged suggestions to guide and refine the narrative structure. For narrative content, the system’s suggestions also played a pivotal role in helping participants develop the flow and focus of their presentations. For instance, during a presentation on the “*Introduction and benefits of yoga*,” the system suggested tags such as “*Audience: Beginners*” and “*Benefits: Mental Clarity*,” which the user incorporated into their narrative structure (P02).

These suggestions helped guide users in tailoring their presentation content to the intended audience and goals, adding more depth to the narrative while reducing the cognitive load of coming up with all the necessary elements from scratch.

Users frequently explored and chose drop-down suggestions. We observed several distinct patterns in how participants interacted with and utilized the suggested tags: In some cases, users directly dragged the suggested tag into a relevant group and then changed its value by selecting an option from the alternatives drop-down widget (see sequences of a red vertical stroke followed by an orange circle ● in Figure 17, center lane).

For example, when the system suggested “cycling: cycling for fitness” for the topic of road riding, one participant dragged it into the narrative group and then refined it by selecting “cycling for environmental benefit” from the dropdown (P06, see Figure 17 around minute 2).

Users frequently overrode system suggestions. Another pattern we frequently observed was that participants often took more exploratory approaches, such as dragging the tag into the group and examining the available dropdown values before manually overriding the tag’s value or label to better suit their vision (see sequences of a red vertical stroke followed by a purple square ■ in Figure 17, center lanes).

Participants requested image suggestions in quick succession. Image suggestions triggered by the system followed different workflows, as the system provided five different images each time, prompting some participants to drag and drop the suggested images into the content sources before triggering the system again for more options. (see sequences of a cyan triangle followed by pink vertical strokes ▼ | in Figure 17, lower lanes)

Users explored the slider but not for steering. Additionally, while participants occasionally explored slider options to see the opposite ends of a spectrum, this was done more for curiosity than for direct steering slide generation. This indicates a potential area for further refinement in how the slider functionality is integrated into the generation process.

In the Slide Steering Overlay, users primarily used “Grounding Tags” for adjustments. Lastly, those participants who used the Slide Steering Overlay to adjust single slides primarily engaged with the tags that the system had generated through its grounding from slide feature (see sequences of a grey diamond in the upper lanes followed by events in the center lanes in Figure 17, upper lanes). For example, when P09 invoked the *Tag Steering Overlay* for adjusting a single slide, the system created multiple tags representing the slide’s current state, one in the visual style group stating “Color: Red and Black Text.” To make adjustments, the user first explored the drop-down widget’s alternative suggestions and then overrode the tag’s value with “blue, red, and white,” resulting in a slide variation with red and white text on a blue background.

6.3 Users’ Perceived Benefits and Challenges for Working with Intent Tags (RQ3)

We clustered the perceived benefits and challenges for users working with intent tags into three themes: *Expressing intents with tags*, *catering users’ presentation needs and workflows*, and *meta-intent elicitation*.

6.3.1 Expressing Intents with Tags. All participants found intent tags advantageous compared to chat-based prompts, appreciating their conciseness and the ease of specifying inputs without needing to write out long sentences: “it’s quite like fun to be able to sort of click and not have to write out a lot [...] I think the benefit of having these labels [over chat-based] is the conciseness. I don’t have that pressure to have to form a sentence.” (P02)

Users liked micro-prompting for its granularity and control. Many participants also appreciated the granularity and control of intent tags while being able to quickly see which presentation-related aspects they can specify: “I really liked that granularity where you could specify, like, this is exactly the font style and I want the kind of theme of it. I like that aspect a lot [...] visually seeing these buckets of information where I can specify different tags. I think that is helpful.” (P07)

Users valued the structure that tag grouping provided. Many participants found the tag groups (e.g., narrative, content source, visual style) intuitive and helpful for sequencing and structuring their process and thinking more explicitly about presentation-related aspects: “I think that narrative is super cool in terms of thinking about the audience [...] it helped me be more explicit about who my audience is in the deck that I’m creating. And it helped me make the right deck faster.” (P11)

Users expected to define their own tag groups. However, other users questioned the necessity of pre-defined groups, suggesting the ability to customize or even work without tag groupings, which might better align with their workflow, as their needs can vary between different slide decks: “But also I would want the ability to just make my own [groups] because I think between different slide decks, it would change” (P08)

Users needed a moment to get used to tags’ flexible terminology. Some participants experienced initial uncertainty around the terminology of intent tags, particularly when classifying and naming tags or choosing appropriate groups: “So I found it slightly intimidating at the beginning because there is some kind of arbitrary way of classifying.” (P09) However, with experimentation and practice, they realized these decisions were not critical as the system does not rely on precise syntax but allows for flexible ways of specifying tags: “But after doing it for a while, it seems like it’s not that important either. Like, this decision of what to put in the title and this kind of things... As far as I’m putting text, I figure that it does something.” (P09)

Users appreciated the preview features. Many appreciated the preview feature as it provided a sense of security and confidence when using the system. Knowing that they could see a preview before finalizing their choices reduced the fear of unexpected outcomes but also helped their imagination: “I thought that the slider was neat. [...] I do appreciate how I’m not just imagining, like on a scale from one to five, how much detail do you want? I don’t know... So, I appreciate that it has a little preview down there. That’s neat.” (P05)

Users expressed a desire for predefined tags. Some participants also noted that starting with a blank tag canvas could be overwhelming, so having a set of predefined tags (such as audience, narrative style, and number of slides) would provide guidance and help users get started: “Let’s say empty tags where there’s maybe already kind of like a label... like audience, but there’s no audience

mentioned yet. But for some people, they don't think about audience before they make a PowerPoint. So, just to be like, hey, who is this for? Is sometimes the most helpful feedback you can give someone. So you can get it here." (P05)

Users expected to rank tags. Some participants also expressed the need for a feature that allows them to rank or prioritize tags based on their importance to the presentation. One participant (P03) mentioned how they tried to emphasize certain points by grouping similar tags together, such as audience and objective, and incorporating system-suggested tags. They also noted that some tags, like colors or themes, were less important to them.

6.3.2 Supporting Diverse Presentation Needs and Workflows. In the interviews, participants also reflected on their general professional slide-creation experience and how IntentTagger could enhance or alleviate their process. Many described the challenges of traditional tools where the manual slide-by-slide process of creating a deck often feels tedious and uninspiring: *"This was a really fun way to create a deck. I think it would appeal to most people that work in PowerPoint, because I don't think most people in PowerPoint creating decks just don't enjoy it unless they do."* (P02)

Users valued the system for overcoming 'blank page' barrier. Many participants found the system particularly helpful in overcoming the initial challenges of starting a presentation, which they identified as one of the hardest parts of any creative project. The ability to quickly generate a basic draft allowed them to move past the initial "blank page" barrier, making it easier to start thinking about additional content and refinements. This was especially valuable for those who often start presentations from scratch, as the system provided a visual outline that lowered the barrier to getting started, allowing them to focus more on editing and refining the content: *"... I feel like the system got the basics going, which then prompted me to think of additional things to add on top."* (P05)

Participants appreciated IntentTagger's flexibility in supporting diverse workflows. Participants appreciated the system's ability to support various workflows, whether by starting with a detailed outline, a brain dump, or a narrative-first approach. Some participants highlighted the system's flexibility in allowing them to begin with content and then refine it iteratively, while others valued the ability to focus on narrative and then move on to visual elements. The system's adaptability to different working styles, such as sequential versus big-picture thinking, was also noted as beneficial.

Participants appreciated the outline editor for structuring ideas and overview. Additionally, the Outline Editor was particularly praised for helping users map out their ideas and providing a clear presentation overview, making it easier to refine individual slides or the entire deck as needed. One participant (P06) also preferred a more sequential workflow, suggesting that instead of showing all tag groups at once, they could focus on one at a time.

Users saw value in tags for tailoring presentations to different purposes. Furthermore, some users mentioned the system's tag-based approach could be useful for creating different versions of presentations tailored to specific audiences or contexts. The ability to easily drag tags in and out allowed for quick adjustments, making it convenient to include or exclude certain content based on the needs of a shorter talk, a quick checkup, or a presentation

adapted for a particular audience: *"Sometimes I need to make different versions of presentations [...] This tag system of dragging things out would be an easy way to include and exclude some things if I needed to."* (P03)

Participants appreciated the fluid integration of deck-wide and slide-specific edits. Participants valued the system's ability to manage both the entire deck and individual slides, appreciating the flexibility to make broad changes across the deck or to focus on specific slides for detailed adjustments. This dual approach was seen as more efficient compared to traditional slide authoring workflows, where changes often have to be made slide-by-slide: *"I think that having this create the whole deck in one shot is super time-saving, and I don't expect it to be hit in one shot. And having the option to click into each slide, and if I like what I do here and want to apply it to the rest... I like the ability to go from everything, update everything, zoom in on one, and then experiment on that one. And then again, update all."* (P02)

Users demanded more control by locking specific design elements. While they appreciated the flexibility and ambiguity offered by the system, participants expressed a desire for more control over certain aspects of their presentations by being able to lock or pin specific slide elements (like background colors, fonts, or other design choices) once they were satisfied with them to prevent them from changing unexpectedly during further generation cycles: *"I do love the slider. I only just don't love that if I slide it one way and I get the color I want that it can subsequently change. I want to be able to anchor it."* (P11)

6.3.3 Meta-Intent Elicitation: Helping Creators Figure Out What They Want and Need. One of the key benefits highlighted by participants was the system's ability to help them discover and refine their intentions when creating presentations. Many mentioned that in their daily work, the process of selecting the right content and design elements was often difficult without clear guidance or inspiration. The system's suggestions helped participants think more critically about their choices and provided new ideas they might not have initially considered. This process of "meta-intent" elicitation allowed users to identify important elements that they may have overlooked or not realized they wanted to include, making it easier to achieve their desired outcome: *"I thought it was cool and interesting because it suggests things that I just didn't think about... And I like these suggestions because I don't always know what are the best style choices for something."* (P03)

Participants appreciated the system guidance's gradual granularity Participants appreciated the system's converging granularity through its suggestions, allowing them to gradually refine their presentations in more specific and thoughtful ways: *"I feel like [IntentTagger] suggests a lot more things that I can control. Versus with [the chat-based system], there are like four prompts and they're like buckets of things you can do rather than the specific things you can tweak. And so for me, this system of these tags feels like I can get more granular towards some kind of vision that I want because I think inherently I'm making some of those choices."* (P12) This guidance made the process more engaging and aligned with their vision while also simplifying the decision-making process by presenting contextually relevant options and ideas to respond to: *"It kind of breaks down those decision points that I'm doing sort of subconsciously in a*

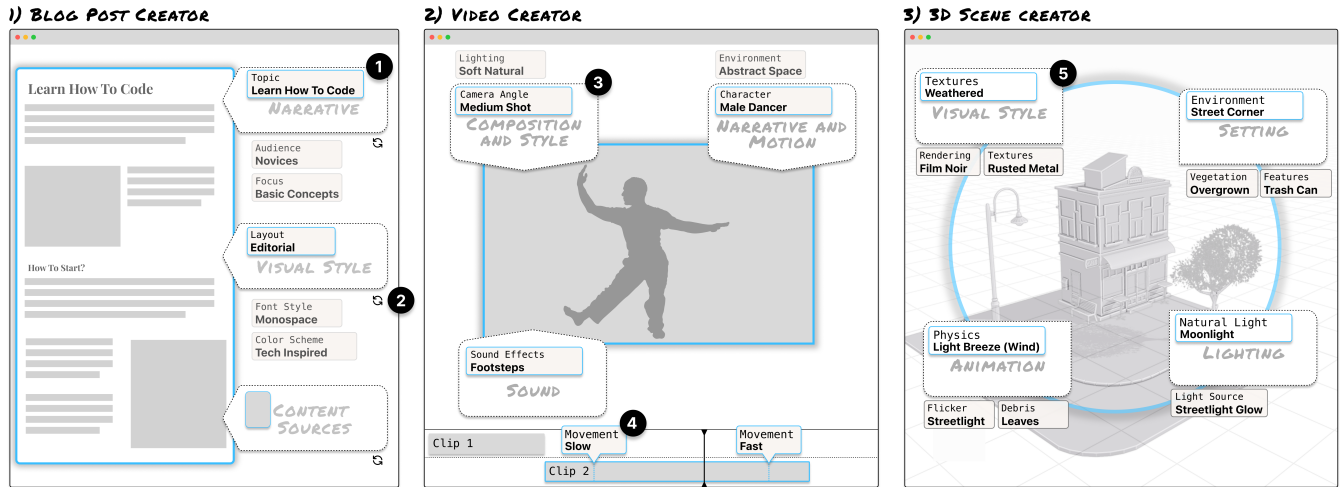


Figure 18: UX interface sketches illustrating alternative use-cases for Intent Tagging-based steering of GenAI content creation; (Left) Blog post creator: (1) Steering of a post generation through intent tags in a side panel; (2) new tag suggestion can be requested per group; (middle) Video creator: Steering of a selected video clip through intent tags; (4) additionally, Keyframe Tags allow to control temporal aspects of the clip generation in the timeline view; (Right) 3D scene creator: (5) Steering of a 3D scene generation through intent tags.

way that is fairly easy. And, like, it suggests a lot of different things [...] it feels like I can get to a decent outcome fairly quickly." (P12)

Participants imagined Intent Tagging to be beneficial for **prototyping presentations and initial brainstorming**. For some participants who reported frequently struggling to organize their thoughts or create structured presentations, the system could also serve as a useful tool for brainstorming and organizing their thoughts. Through the drag-and-drop interface, participants imagined to quickly include or exclude elements, allowing them to organize and prototype their presentations more intuitively and efficiently: "I can organize my thoughts down and drag things in I want to include, drag things out that I want to exclude." (P03)

Some users felt overwhelmed by visual clutter and suggestions. However, while the majority of participants found the amount and visual arrangement of the system-suggested tags helpful, a few noted that too many suggestions could feel visually cluttered and overwhelming. Some expressed a preference for a more simplified or list-based representation, allowing them to focus more easily on the suggestions without feeling overwhelmed by the options: "...so a bit more visual hierarchy, maybe less overloaded with suggestions." (P01)

Participants praised the slider for stimulating their critical thinking. Additionally, features such as the opposite slider helped participants think more deeply about the intention behind their choices. By offering a way to explore different dimensions of a tag or design element, participants were able to visualize various possibilities and fine-tune their decisions in a way that aligned with their goals and encouraged reflection on the impact of their choices: "I really like the slider idea of like... 'oh, you have this idea, but then you want to see what the opposite end is,' and you can put a slider on

that value [...] it just makes you think more about your intention of what you're trying to do." (P04)

7 Intent Tagging Interactions Beyond Slide Deck Creation

To illustrate how intent tagging could be utilized to facilitate human-GenAI co-creation interactions across diverse rich content creation tasks beyond slide deck creation, we explored alternative **user experience (UX) interface scenarios** for rich content creation applications. Some of these were inspired by participants' speculations during the interviews on how IntentTagger could be helpful in applications like Excel for complex data visualizations, educational materials, and tasks such as creating videos or book reports.

In a design sprint, we sketched out three UX interface scenarios for intent tag-based steering of GenAI-driven creation of blog posts, videos, and 3D scenes (Figure 18). We defined different tag groups for each application, matching the creation task's high-level categories. We also explored how to possibly expand the "tag vocabulary" and their functionalities by, for example, envisioning *Keyframe Tags* to allow control over temporal aspects of clips in the video creator (Figure 18-4). Overall, while only sketches, these explorations aim to highlight the versatility of intent tagging for steering rich content generation across a range of applications and interfaces to be further explored in future research.

8 Discussion and Design Considerations

In the following sections, we discuss our explorations' learnings from the user study and UX scenarios, reflecting on *Intent Tagging as an enabler of non-linear GenAI workflows* and its role in *promoting Human-AI Co-creation and reflection*. Further, we will discuss the challenges of *blending steering interactions alongside manual*

content editing and, lastly, reflect on *trade-offs between UI customization and user distractions*. Each section discusses challenges, design considerations, and opportunities for future work.

8.1 Intent Tagging Enables Non-Linear GenAI Workflows

Our findings demonstrate how IntentTagger was able to facilitate flexible, non-linear, and non-destructive workflows with high degrees of freedom: Once users became comfortable with the tool, they were able to seamlessly switch between the Deck Steering Canvas, Slide Overlay, and Outline Editor and steering the GenAI system without having to follow a rigid sequence (see 6.2). However, this non-linear steerability comes with a cost: While participants valued the system's flexibility, they also expressed frustration about a need for greater control by "locking" or "pinning" specific content or design elements, such as background colors and fonts, once satisfied to prevent unintended changes during further generation cycles (see 6.5). While it is a powerful concept to enable multi-modal and multi-directional steering of GenAI content generation between tag, outline, slide, and deck levels, we currently lack interfaces allowing users to gradually lock elements across such dimensions. As a starting point, future research in this direction could be partially informed by design space exploration and parametric design systems that support users navigating and gradually reducing the degree of freedom across large option sets [55, 99].

In terms of user experience, Intent Tagging presents a novel way of steering GenAI systems and creating slide decks that are new to users. Our findings showed that while all participants appreciated the flexibility to specify tags with different levels of specificity, many were initially unsure about the tags' correct terminology and group associations (see 6.2.2). This uncertainty posed initial challenges, which participants quickly overcame after experiencing the systems' flexibility in interpreting their intent after the first generations. To better support user onboarding, we suggest that future intent tagging-based interfaces should more explicitly scaffold that learning process, for example, by partially pre-defining common tag labels (such as "*audience*" or "*length*") or offering optional "tag templates" for specific content creation scenarios to help users better get started and learn intent tagging.

8.2 Intent Tagging Promotes Human-AI Co-creation and Reflection-In-Action

Figuring out one's intentions and requirements is often the hardest part of complex tasks like slide deck creation. Generally, *Schön* describes this process as "reflection-in-action" [78], where creators interact with the material at hand while the material "talks back" to them, creating an iterative process of acting and reflection. When intent tagging, the tags on the canvas talk back to the user, whether self-generated or system-suggested, helping them (along with the outline and slides) tackle the cognitively demanding task of clarifying their intent. Furthermore, the more task-relevant tags a user specifies, the more accurately the system can provide contextually tailored suggestions, better aligning with the user's objectives. As users add more tags, they convey a fuller picture to the GenAI system, which in turn enhances the quality and precision of the generated content. The system and user engage in a close feedback

loop, allowing users to refine their intentions while the AI system responds dynamically—like connecting a mind map to a GenAI content creation system. This process relates closely to previously framed "enactive" models of Human-AI co-creation [22], relating to the theory that cognition arises from improvised interactions shaped by feedback from the surrounding environment.

In this version of IntentTagger, we decided that users must explicitly request tag and image suggestions from the system. On the one hand, this approach ensures persistency between requests, allowing users to explore suggested tags without the risk of suggestions disappearing without their consent. On the other hand, tag suggestions also offer an exciting possibility of being combined with mixed-initiative interactions in the future, where the system can automatically initiate tag suggestions in response to the user's active tags [42].

In terms of user-system communication, in this study, we focused on analyzing users' general interactions with system-generated tags (see 6.2.3). Future work could delve into a deeper semantic analysis, exploring how users expressed their intent and what specific tags the system suggested in response. Such insights could further improve the effectiveness of tag-based human-AI co-creation systems.

Furthermore, while many participants experimented with the *Opposite Slider widget*, no one used it to steer slide generation (see 6.2.3). However, several participants voiced the clear value of this feature for prompting reflection on their design choices (see 6.6). This suggests that, while not used directly for steering generation, the slider still serves as a valuable tool for encouraging and steering deeper thinking. This bears interesting further implications for designing future GenAI steering interfaces, which might consider integrating steering controls alongside "intent exploration" UI to promote reflection independent of content generation.

Our findings also revealed that users often got inspired by system suggestions but then manually overrode the tag's value field to better align with their vision. This interaction pattern hints at an effective combination of guided intent exploration through system-generated options alongside the flexibility for users to override suggestions if desired. Adding to this observation, a recent study in the context of AI-assisted writing found that participants, despite the many integrated AI suggestions, felt ownership over the content "*because of the numerous authorial decisions*" they made [83]. Together, such findings suggest that it is critical for effective human-AI co-creation to enable users to manually edit and override system suggestions to foster a sense of ownership and control.

Our research indicates that user interactions with IntentTagger's LLM-generated *tag suggestions*, *tag grounding*, and *pre-generated real-time previews* can help align user expectations with GenAI systems' capabilities and recognize their limitations (see 6.4). While some suggested tags referred to aspects beyond the slide generation capabilities (such as icons or background textures), interestingly, users were not significantly frustrated by that. Instead, they found the suggestions helpful for improving their presentations, regardless of whether the system could fully execute them. This tolerance was likely influenced by the pre-generated real-time tooltip previews, which helped users quickly grasp the system's limitations without frustration, as they had not invested much time in these explorations.

Adding to previous work on surfacing affordances and providing feed-forward mechanisms [10, 89, 91], we promote dynamically **pre-generating** slide previews (or other content) on a granular level (such as only alternating a slide’s font or text format) to support users’ intent exploration and foster greater tolerance when GenAI falls short by reducing emotional investment in misaligned expectations. Additionally, *tag grounding*—such as the grounding from text prompt or slide features—might also improve alignment between user expectations and system capabilities by helping users quickly understand how the system interprets their inputs. Lastly, presenting distinct differences between versions of generated content may further enhance reflection-in-action during the creation process, building on insights from Drucker et al. [25]. All these mechanisms show promising potential for enhancing user alignment in GenAI workflows.

8.3 Blending Steering Interactions and Manual Content Editing

Our prototype system is designed to probe the concept of intent tagging-based interactions for steering slide deck generation and, therefore, does not support rich content editing that full-fledged applications such as PowerPoint offer. In the study, the system allowed participants to create slide decks with GenAI support with more control and satisfaction than with the chat-based system. However, going further, we envision that IntentTagger would also integrate rich content editing capabilities of slides in addition to AI-driven generation. Enabling users to interweave manual and AI-driven editing activities opens up many interesting questions relating to the orchestration of the hand-off between user and system. For example, devising mechanisms for identifying situations in which steering GenAI actually becomes more tedious than directly editing content and surfacing cues to users nudging them towards direct editing is an intriguing direction to explore.

8.4 Trade-offs between UI Customization and UI Management

Lastly, we reflect on requests we gathered from participants to enable them to customize IntentTagger more. A salient example is the requests from a few participants for enabling customizable groupings of tags rather than the three fixed groups we offered (see 6.4). While relatively straightforward to achieve, such added capability may come as a tradeoff between user control of the interface and its management.

While giving users the ability to create their own tag groups’ structure would allow for greater flexibility—especially for catering to users’ specific needs and creation process—it would also add an additional cognitive layer by requiring users to reflect on what such process is and how to best articulate it in the interface. This leads to the classic question of identifying which interface aspects are best given to the users to shape and which to leave in the hands of design and system architects. Probing this delicate balance between supporting flexibility and minimizing effort, enabling rich interactions, and making them simple to use in the context of generative AI paves the way for future research in our field.

8.5 Limitations of the Evaluation

We highlight three limitations of this work: Firstly, all participants were recruited from the same large technology company, and while they represent a diverse set of job titles and backgrounds, it is possible that participants from other companies or from non-corporate environments might provide additional insights not captured in the current study. Secondly, the study’s lab setting, limited task times, and constrained number of slides allowed us to study Intent Tagging in a controlled and concise format. However, future work is needed to study the impacts of Intent Tags on creation processes *in-vivo* and over *longer periods of time* to gain further insights into real-life content creation scenarios. Lastly, we deliberately limited our comparison of IntentTagger with the widespread commercial Copilot and Designer features in PowerPoint. Future comparisons with other tools and interfaces may surface a broader range of benefits and tradeoffs.

9 Conclusion

GenAI offers immense potential for augmenting content creation, but challenges around intent elicitation and alignment, prompt formulation, and workflow flexibility impede effective human-AI co-creation. To address these issues, we propose intent tagging: graphical micro-prompting interactions for supporting granular and non-linear workflows with GenAI systems in the context of slide deck creation. To explore its benefits and challenges, we developed IntentTagger, an intent tagging-based LLM-driven system allowing users to iteratively create and modify slide deck presentations, and conducted a user study. Our findings revealed that users preferred intent tag-based interactions over chat- and gallery-based systems, valuing the system’s ability to support non-linear workflows, flexible intent expression, and integrated suggestions that helped them clarify their goals and think through the slide creation tasks. Based on these findings, we discuss design considerations for integrating intent tagging into GenAI-assisted slide authoring. Although our system and study primarily focus on slide creation, intent tags seem to present a design pattern that researchers could explore further in other AI-assisted workflows and scenarios, thereby offering designers a new interaction technique for HCI+AI applications.

Acknowledgments

We thank all study participants, as well as Anna Offenwanger and Zheng Ning, for their support and feedback throughout the project and the reviewers for their constructive feedback on the paper. This work was supported in part by the National Science Foundation under Grant No. 2118924.

References

- [1] Adobe. 2024. Adobe Firefly. <https://www.adobe.com/products/firefly.html>.
- [2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland UK, 1–13. <https://doi.org/10.1145/3290605.3300233>
- [3] Robert Rene Henri Anholt. 2006. *Dazzle ‘em with Style: The Art of Oral Scientific Presentation* (2nd ed ed.). Elsevier Academic Press, Philadelphia.
- [4] ATLAS.ti. 2024. ATLAS.ti. <https://atlasti.com>.

- [5] Alex Baker and André van der Hoek. 2010. Ideas, Subjects, and Cycles as Lenses for Understanding the Software Design Process. *Design Studies* 31, 6 (Nov. 2010), 590–613. <https://doi.org/10.1016/j.destud.2010.09.008>
- [6] Eric P.S. Baumer and Bill Tomlinson. 2011. Comparing Activity Theory with Distributed Cognition for Video Analysis: Beyond "Kicking the Tires". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 133–142. <https://doi.org/10.1145/1978942.1978962>
- [7] James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, Wesam Manassra, Prafulla Dhariwal, Casey Chu, Yunxin Jiao, and Aditya Ramesh. 2023. Improving Image Generation with Better Captions.
- [8] BlockNote. 2024. BlockNote - Javascript Block-Based React Rich Text Editor. <https://www.blocknotejs.org>.
- [9] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kavin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshthe Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kudipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suver Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Nieves, Hamed Nilforoshan, Julian Nyarko, Giray Ogun, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. 2022. On the Opportunities and Risks of Foundation Models. <https://doi.org/10.48550/arXiv.2108.07258> arXiv:2108.07258 [cs]
- [10] Jeremy Boy, Louis Eveillard, Françoise Detienne, and Jean-Daniel Fekete. 2015. Suggested interactivity: Seeking perceived affordances for information visualization. *IEEE transactions on visualization and computer graphics* 22, 1 (2015), 639–648.
- [11] Stephen Brade, Bryan Wang, Mauricio Sousa, Sageev Oore, and Tovi Grossman. 2023. Promptify: Text-to-Image Generation through Interactive Prompt Exploration with Large Language Models. arXiv:2304.09337 [cs]
- [12] Virginia Braun and Victoria Clarke. 2019. Reflecting on Reflexive Thematic Analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (Aug. 2019), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- [13] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models Are Few-Shot Learners. <https://arxiv.org/abs/2005.14165v4>.
- [14] Bill Buxton. 2007. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann, Burlington. <https://doi.org/10.1016/B978-0-12-374037-3.X5043-3>
- [15] Wanling Cai and Li Chen. 2020. Predicting User Intent and Satisfaction with Dialogue-based Conversational Recommendations. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '20)*. Association for Computing Machinery, New York, NY, USA, 33–42. <https://doi.org/10.1145/3340631.3394856>
- [16] Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Michael Petrov, Heidy Khlaaf, Girish Sastry, Pamela Mishkin, Brooke Chan, Scott Gray, Nick Ryder, Mikhail Pavlov, Alethea Power, Lukasz Kaiser, Mohammad Bavarian, Clemens Winter, Philippe Tillet, Felipe Petroski Such, Dave Cummings, Matthias Plappert, Fotios Chantzis, Elizabeth Barnes, Ariel Herbert-Voss, William Hebgren Guss, Alex Nichol, Alex Paino, Nikolas Tezak, Jie Tang, Igor Babuschkin, Suchir Balaji, Shantanu Jain, William Saunders, Christopher Hesse, Andrew N. Carr, Jan Leike, Josh Achiam, Vedant Misra, Evan Morikawa, Alec Radford, Matthew Knight, Miles Brundage, Mira Murati, Katie Mayer, Peter Welinder, Bob McGrew, Dario Amodei, Sam McCandlish, Ilya Sutskever, and Wojciech Zaremba. 2021. Evaluating Large Language Models Trained on Code. arXiv:2107.03374 [cs]
- [17] Leonardo Christino, Taylor Hill, and Fernando Paulovich. 2022. Knowledge-Decks: Automatically Generating Presentation Slide Decks of Visual Analytics Knowledge Discovery Applications. <https://doi.org/10.48550/arXiv.2212.01469> arXiv:2212.01469 [cs]
- [18] John Joon Young Chung and Eytan Adar. 2023. PromptPaint: Steering Text-to-Image Generation Through Paint Medium-like Interactions. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. <https://doi.org/10.1145/3586183.3606777>
- [19] John Joon Young Chung, Woosok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories with Generative Pretrained Language Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–19. <https://doi.org/10.1145/3491102.3501819>
- [20] Alan Cooper, Robert Reimann, Dave Cronin, and Alan Cooper. 2014. *About Face: The Essentials of Interaction Design* (fourth edition ed.). John Wiley and Sons, Indianapolis, IN.
- [21] Maria João Costa, Hugo Amaro, and Hugo Gonçalves Oliveira. 2023. SmartEDU: Accelerating Slide Deck Production with Natural Language Processing. In *Lecture Notes in Computer Science*. Springer Nature Switzerland, 109–123.
- [22] Nicholas Davis, Chih-Pin Hsiao, Yanna Popova, and Brian Magerko. 2015. An Enactive Model of Creativity for Computational Collaboration and Co-creation. In *Creativity in the Digital Age*, Nelson Zagalo and Pedro Branco (Eds.). Springer London, London, 109–133. https://doi.org/10.1007/978-1-4471-6681-8_7
- [23] Tomas Dorta, Annemarie Lesage, Edgar Pérez, and J. Bastien. 2010. Signs of Collaborative Ideation and the Hybrid Ideation Space. 199–206. https://doi.org/10.1007/978-0-85729-224-7_26
- [24] Graham Dove, Nicolai Brodersen Hansen, and Kim Halskov. 2016. An Argument For Design Space Reflection. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. Association for Computing Machinery, New York, NY, USA, 1–10. <https://doi.org/10.1145/2971485.2971528>
- [25] Steven M. Drucker, Georg Petschnigg, and Maneesh Agrawala. 2006. Comparing and Managing Multiple Versions of Slide Presentations. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*. ACM, Montreux Switzerland, 47–56. <https://doi.org/10.1145/1166253.1166263>
- [26] Darren Edge, Joan Savage, and Koji Yatani. 2013. HyperSlides: Dynamic Presentation Prototyping. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris France, 671–680. <https://doi.org/10.1145/2470654.2470749>
- [27] John H. Flavell. 1979. Metacognition and Cognitive Monitoring: A New Area of Cognitive–Developmental Inquiry. *American Psychologist* 34, 10 (1979), 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>
- [28] React Flow. 2024. React Flow. <https://reactflow.dev/>.
- [29] FormidableLabs. 2024. FormidableLabs/Spectacle. Nearform Commerce.
- [30] Tsu-Jui Fu, William Yang Wang, Daniel McDuff, and Yale Song. 2022. DOC2PPT: Automatic Presentation Slides Generation from Scientific Documents. *Proceedings of the AAAI Conference on Artificial Intelligence* 36, 1 (June 2022), 634–642.
- [31] Steve Garner and Deana McDonagh-Philp. 2001. Problem Interpretation and Resolution via Visual Stimuli: The Use of 'Mood Boards' in Design Education. *Journal of Art & Design Education* 20, 1 (Feb. 2001), 57–64. <https://doi.org/10.1111/1468-5949.00250>
- [32] GitHub. 2024. GitHub Copilot. <https://github.com/features/copilot>.
- [33] Frederic Gmeiner, Humphrey Yang, Lining Yao, Kenneth Holstein, and Nikolas Martelaro. 2023. Exploring Challenges and Opportunities to Support Designers in Learning to Co-create with AI-based Manufacturing Design Tools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–20. <https://doi.org/10.1145/3544548.3580999>
- [34] Grammarly. 2024. Grammarly. <https://grammarly.com/>.
- [35] Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P. Kingma, Ben Poole, Mohammad Norouzi, David J. Fleet, and Tim Salimans. 2022. Imagen Video: High Definition Video Generation with Diffusion Models. <https://arxiv.org/abs/2210.02303v1>.
- [36] Peiling Jiang, Jude Rayan, Steven P. Dow, and Haijun Xia. 2023. Graphologue: Exploring Large Language Model Responses with Interactive Diagrams. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–20. <https://doi.org/10.1145/3586183.3606737>
- [37] G. Klein, B. Moon, and R.R. Hoffman. 2006. Making Sense of Sensemaking 1: Alternative Perspectives. *IEEE Intelligent Systems* 21, 4 (July 2006), 70–73. <https://doi.org/10.1109/MIS.2006.75>
- [38] Jon Kolko. 2010. Sensemaking and Framing: A Theoretical Reflection on Perspective in Design Synthesis. (2010).
- [39] Max Kreminski and John Joon Young Chung. 2024. Intent Elicitation in Mixed-Initiative Co-Creativity. In *IUI Workshops*.
- [40] Max Kreminski, Melanie Dickinson, Joseph Osborn, Adam Summerville, Michael Mateas, and Noah Wardrip-Fruin. 2020. Germinate: A Mixed-Initiative Casual Creator for Rhetorical Games. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* 16, 1 (Oct. 2020), 102–108.

- <https://doi.org/10.1609/aiide.v16i1.7417>
- [41] Leonhard Lichtschlag, Thorsten Karrer, and Jan Borchers. 2009. Fly: A Tool to Author Planar Presentations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Boston MA USA, 547–556. <https://doi.org/10.1145/1518701.1518786>
 - [42] Zhiyu Lin, Upol Ehsan, Rohan Agarwal, Samihan Dani, Vidushi Vashishth, and Mark Riedl. 2023. Beyond Prompts: Exploring the Design Space of Mixed-Initiative Co-Creativity Systems. <https://doi.org/10.48550/arXiv.2305.07465>
 - [43] Michael Xieyang Liu, Advait Sarkar, Carina Negreanu, Benjamin Zorn, Jack Williams, Neil Toronto, and Andrew D. Gordon. 2023. “What It Wants Me To Say”: Bridging the Abstraction Gap Between End-User Programmers and Code-Generating Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–31. <https://doi.org/10.1145/3544548.3580817>
 - [44] Michael Xieyang Liu, Tongshuang Wu, Tianying Chen, Franklin Mingzhe Li, Aniket Kittur, and Brad A. Myers. 2024. Selenite: Scaffolding Online Sensemaking with Comprehensive Overviews Elicited from Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–26. <https://doi.org/10.1145/3613904.3642149> arXiv:2310.02161 [cs]
 - [45] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021. Pre-Train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. arXiv:2107.13586 [cs]
 - [46] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. arXiv:2204.09007 [cs]
 - [47] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J. Cai. 2020. Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. <https://doi.org/10.1145/3313831.3376739>
 - [48] M. Sravanthi, C. R. Chowdary, and P. S. Kumar. 2009. SlidesGen: Automatic Generation of Presentation Slides for a Technical Paper Using Summarization. *The Florida AI Research Society* (2009).
 - [49] Atefeh Mahdavi Goloujeh, Anne Sullivan, and Brian Magerko. 2024. Is It AI or Is It Me? Understanding Users’ Prompt Journey with Text-to-Image Generative AI Tools. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI ’24)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3613904.3642861>
 - [50] Chris Martens and Matthew A. Hammer. 2017. Languages of Play: Towards Semantic Foundations for Game Interfaces. In *Proceedings of the 12th International Conference on the Foundations of Digital Games (FDG ’17)*. Association for Computing Machinery, New York, NY, USA, 1–10. <https://doi.org/10.1145/3102071.3102096>
 - [51] Marvin. 2024. Marvin. <https://heymarvin.com/>.
 - [52] Utiyama Masao and Hasida Kōiti. 1999. Automatic Slide Presentation from Semantically Annotated Documents. In *Proceedings of the Workshop on Coreference and Its Applications - CorefApp ’99*. Association for Computational Linguistics, College Park, Maryland, 25. <https://doi.org/10.3115/1608810.1608816>
 - [53] Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. 2023. Statslator: Interactive Translation of NHST and Estimation Statistics Reporting Styles in Scientific Documents. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–14. <https://doi.org/10.1145/3586183.3606762>
 - [54] Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. 2024. Direct-GPT: A Direct Manipulation Interface to Interact with Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI ’24)*. Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3613904.3642462>
 - [55] Justin Matejka, Michael Glueck, Erin Bradner, Ali Hashemi, Tovi Grossman, and George Fitzmaurice. 2018. Dream Lens: Exploration and Visualization of Large-Scale Generative Design Datasets. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–12. <https://doi.org/10.1145/3173574.3173943>
 - [56] Microsoft. 2024. Bing Search API. <https://www.microsoft.com/en-us/bing/apis/bing-web-search-api>.
 - [57] Microsoft. 2024. PowerPoint. <https://www.microsoft.com/en-us/microsoft-365/powerpoint>.
 - [58] Microsoft. 2024. PowerPoint Copilot. <https://support.microsoft.com/en-us/office/create-a-new-presentation-with-copilot-in-powerpoint-3222ee03-f5a4-4d27-8642-9c387ab4854d>.
 - [59] Microsoft. 2024. PowerPoint Designer. <https://support.microsoft.com/en-us/office/create-professional-slide-layouts-with-designer-53c77d7b-dc40-45c2-b684-814155eac0617>.
 - [60] Miro. 2024. Miro. <https://miro.com/>.
 - [61] Ine Mols, Elise Van Den Hoven, and Berry Eggen. 2016. Informing Design for Reflection: An Overview of Current Everyday Practices. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. ACM, Gothenburg Sweden, 1–10. <https://doi.org/10.1145/2971485.2971494>
 - [62] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 2018. I Lead, You Help but Only with Enough Details: Understanding User Experience of Co-Creation with Artificial Intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI ’18)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3173574.3174223>
 - [63] OpenAI. 2024. OpenAI API. <https://platform.openai.com>.
 - [64] Jieun Park, Sangyeon Kim, and Sangwon Lee. 2023. A User Preference and Intent Extraction Framework for Explainable Conversational Recommender Systems. In *Companion Proceedings of the 2023 ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS ’23 Companion)*. Association for Computing Machinery, New York, NY, USA, 16–23. <https://doi.org/10.1145/3596454.3597178>
 - [65] Savvas Petridis, Hujung Valentina Shin, and Lydia B Chilton. 2021. SymbolFinder: Brainstorming Diverse Symbols Using Local Semantic Networks. In *The 34th Annual ACM Symposium on User Interface Software and Technology (UIST ’21)*. Association for Computing Machinery, New York, NY, USA, 385–399. <https://doi.org/10.1145/3472749.3474757>
 - [66] Peter Piroli and Stuart Card. 2005. The Sensemaking Process and Leverage Points for Analyst Technology as Identified through Cognitive Task Analysis.
 - [67] Chen Qu, Liu Yang, W. Bruce Croft, Yongfeng Zhang, Johanne R. Trippas, and Minghui Qiu. 2019. User Intent Prediction in Information-seeking Conversations. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval (CHIIR ’19)*. Association for Computing Machinery, New York, NY, USA, 25–33. <https://doi.org/10.1145/3295750.3298924>
 - [68] R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria.
 - [69] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. <https://arxiv.org/abs/2204.06125v1>.
 - [70] React.js. 2024. React.js. <https://react.dev/>.
 - [71] Garr Reynolds and Guy Kawasaki. 2020. *Presentation Zen: Simple Ideas on Presentation Design and Delivery* (3rd edition ed.). New Riders, San Francisco, CA.
 - [72] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. <https://doi.org/10.48550/arXiv.2112.10752> arXiv:2112.10752 [cs]
 - [73] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Rapha Gontijo Lopes, Tim Salimans, Jonathan Ho, David J. Fleet, and Mohammad Norouzi. 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. <https://arxiv.org/abs/2205.11487v1>.
 - [74] Harshita Sahjiwani. 2022. Adaptive Dialogue Management for Conversational Information Elicitation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR ’22)*. Association for Computing Machinery, New York, NY, USA, 3495. <https://doi.org/10.1145/3477495.3531684>
 - [75] Advait Sarkar. 2023. Exploring Perspectives on the Impact of Artificial Intelligence on the Creativity of Knowledge Work: Beyond Mechanised Plagiarism and Stochastic Parrots. <https://arxiv.org/abs/2307.10751v1>. <https://doi.org/10.1145/3596671.3597650>
 - [76] Wout Schellaert, Fernando Martínez-Plumed, Karina Vold, John Burden, Pablo A. M. Casares, Bao Sheng Loe, Roi Reichart, Sean Ó hÉigeartaigh, Anna Korhonen, and José Hernández-Orallo. 2023. Your Prompt Is My Command: On Assessing the Human-Centred Generality of Multimodal Models. *Journal of Artificial Intelligence Research* 77 (June 2023), 377–394. <https://doi.org/10.1613/jair.1.14157>
 - [77] Johannes Schneider. 2022. Foundation Models in Brief: A Historical, Socio-Technical Focus. (2022). <https://doi.org/10.48550/ARXIV.2212.08967>
 - [78] Donald A. Schön. 1983. *The Reflective Practitioner: How Professionals Think in Action*. Basic Books, New York.
 - [79] Donald A. Schön. 1992. Designing as reflective conversation with the materials of a design situation. *Know-Based Syst.* 5, 1 (mar 1992), 3–14. [https://doi.org/10.1016/0950-7051\(92\)90020-G](https://doi.org/10.1016/0950-7051(92)90020-G)
 - [80] Moushumi Sharmin and Brian P. Bailey. 2011. “I Reflect to Improve My Design”: Investigating the Role and Process of Reflection in Creative Design. In *Proceedings of the 8th ACM Conference on Creativity and Cognition (C&C ’11)*. Association for Computing Machinery, New York, NY, USA, 389–390. <https://doi.org/10.1145/2069618.2069710>
 - [81] Taylor Shin, Yasaman Razeghi, Robert L. Logan IV, Eric Wallace, and Sameer Singh. 2020. AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 4222–4235. <https://doi.org/10.18653/v1/2020.emnlp-main.346>
 - [82] Ben Shneiderman. 2007. Creativity support tools: accelerating discovery and innovation. *Commun. ACM* 50, 12 (Dec. 2007), 20–32. <https://doi.org/10.1145/1323688.1323689>

- [83] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2023. Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence. *ACM Trans. Comput.-Hum. Interact.* 30, 5 (Sept. 2023), 68:1–68:57. <https://doi.org/10.1145/3511599>
- [84] Hari Subramonyam, Roy Pea, Christopher Pondoc, Maneesh Agrawala, and Colleen Seifert. 2024. Bridging the Gulf of Envisioning: Cognitive Challenges in Prompt Based Interactions with LLMs. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–19. <https://doi.org/10.1145/3613904.3642754>
- [85] Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. 2024. Luminate: Structured Generation and Exploration of Design Space with Large Language Models for Human-AI Co-Creation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–26. <https://doi.org/10.1145/3613904.3642400> arXiv:2310.12953 [cs]
- [86] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–18. <https://doi.org/10.1145/3586183.3606756>
- [87] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2023. The Metacognitive Demands and Opportunities of Generative AI. <https://doi.org/10.48550/arXiv.2312.10893> arXiv:2312.10893 [cs]
- [88] Michael Terry, Chinmay Kulkarni, Martin Wattenberg, Lucas Dixon, and Meredith Ringel Morris. 2023. AI Alignment in the Design of Interactive AI: Specification Alignment, Process Alignment, and Evaluation Support. (2023). <https://doi.org/10.48550/ARXIV.2311.00710>
- [89] Michael Terry and Elizabeth D. Mynatt. 2002. Recognizing creative needs in user interface design. In *Proceedings of the 4th Conference on Creativity & Cognition (Loughborough, UK) (C&C '02)*. Association for Computing Machinery, New York, NY, USA, 38–44. <https://doi.org/10.1145/581710.581718>
- [90] Priyan Vaithilingam, Ian Arawjo, and Elena L. Glassman. 2024. Imagining a Future of Designing with AI: Dynamic Grounding, Constructive Negotiation, and Sustainable Motivation. <https://doi.org/10.48550/arXiv.2402.07342> arXiv:2402.07342 [cs]
- [91] Jo Vermeulen, Kris Luyten, Elise van den Hoven, and Karin Coninx. 2013. Crossing the bridge over Norman's Gulf of Execution: revealing feedforward's true identity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1931–1940.
- [92] Fengjie Wang, Yanna Lin, Leni Yang, Haotian Li, Mingyang Gu, Min Zhu, and Huamin Qu. 2024. OutlineSpark: Igniting AI-powered Presentation Slides Creation from Computational Notebooks through Outlines. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3613904.3642865>
- [93] Fengjie Wang, Xuye Liu, Ou Jing Liu, Ali Neshati, Tengfei Ma, Min Zhu, and Jian Zhao. 2023. Slide4N: Creating Presentation Slides from Computational Notebooks with Human-AI Collaboration. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–18. <https://doi.org/10.1145/3544548.3580753>
- [94] Michael Williamson. 2024. Mwilliamson/Mammoth.js.
- [95] Thomas Winters and Kory W. Mathewson. 2019. Automatically Generating Engaging Presentation Slide Decks. In *Lecture Notes in Computer Science*. Springer International Publishing, 127–141.
- [96] Xuyong Yang, Tao Mei, Ying-Qing Xu, Yong Rui, and Shipeng Li. 2016. Automatic Generation of Visual-Textual Presentation Layout. *ACM Transactions on Multimedia Computing, Communications, and Applications* 12, 2 (Feb. 2016), 33:1–33:22. <https://doi.org/10.1145/2818709>
- [97] JoAnne Yates and Wanda Orlikowski. 2007. The PowerPoint Presentation and Its Corollaries: How Genres Shape Communicative Action in Organizations. In *Communicative Practices in Workplaces and the Professions: Cultural Perspectives on the Regulation of Discourse and Organizations*, Mark Zachry and V Thralls (Eds.). Baywood Publishing Company, Inc. <https://doi.org/10.2190/CPIC4>
- [98] Catherine Yeh, Gonzalo Ramos, Rachel Ng, Andy Huntington, and Richard Banks. 2024. GhostWriter: Augmenting Collaborative Human-AI Writing Experiences Through Personalization and Agency. arXiv:2402.08855 [cs]
- [99] Loutfouz Zaman, Wolfgang Stuerzlinger, Christian Neugebauer, Rob Woodbury, Maher Elkhaldi, Naghmi Shireen, and Michael Terry. 2015. GEM-NI: A System for Creating and Managing Alternatives In Generative Design. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 1201–1210. <https://doi.org/10.1145/2702123.2702398>
- [100] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–21. <https://doi.org/10.1145/3544548.3581388>
- [101] Edward Zanders and Lindsay MacLeod. 2018. *Presentation Skills for Scientists: A Practical Guide* (2 ed.). Cambridge University Press. <https://doi.org/10.1017/9781108567640>
- [102] Yaqian Zhu and John Kolassa. 2018. Assessing and Comparing the Accuracy of Various Bootstrap Methods. *Communications in Statistics - Simulation and Computation* 47, 8 (Sept. 2018), 2436–2453. <https://doi.org/10.1080/03610918.2017.1348516>

A Additional Materials

Table 3: Overview of study participants.

ID	Age Range (Years)	Gender	Job Title	Frequency of PowerPoint Usage
P01	18-29	Female	Senior Product Designer	Multiple times per week
P02	30-39	Female	Senior Product Manager	Every day
P03	18-29	Female	Research Intern (PhD)	Multiple times per month
P04	18-29	Male	Senior Human Factors Engineer	Every day
P05	30-39	Female	Senior Designer	Multiple times per week
P06	40-49	Female	Senior User Researcher	Multiple times per week
P07	30-39	Male	Senior Product Designer	Every day
P08	18-29	Female	Content Designer	Multiple times per month
P09	30-39	Male	Research Intern (PhD)	Multiple times per month
P10	18-29	Non-binary	Research Intern (PhD)	Multiple times per year
P11	30-39	Female	Design Program Manager	Multiple times per week
P12	40-49	Female	UX Researcher	Multiple times per month

A.1 Further System Implementation Details

A.1.1 Tag Suggestion Mechanism. To generate intent tag suggestions based on a user's existing tags on the slide deck steering canvas, we construct a call to GPT comprising the static `SYSTEM_PROMPT` and the dynamic `USER_CONTEXT` below. `USER_CONTEXT` is a string dynamically constructed from all active intent tags [attribute:value] for each tag group (Narrative, Visual Style, Content Sources).

SYSTEM_PROMPT (static):

You are an assistant to help users author slide presentations in PowerPoint.

You are helping the user develop the narrative and visual style and extracting content from source documents relevant to the presentation.

You will receive instructions to help with one of these three buckets: Narrative, Visual Style, and Content Sources.

The user will already have descriptions, tags, and media elements that describe different aspects of the presentation.

Your task is to suggest additional keywords and concepts for each of these buckets.

The user will prompt you in the format:

****Bucket**** attribute1:value1, attribute2:value2, ...

A bucket might be empty, in which case you should suggest concepts based on the other buckets.

A value might not be associated with an attribute.

For each user prompt, suggest 7 concepts for each of the requested buckets.

A concept is a keyword or phrase that is relevant to the user's presentation. A concept should be a single word or a short phrase. Each concept should have an attribute and value.

Return your response in JSON format. Only return JSON format without any additional Markdown wrapper code blocks.

Here is an example of the JSON return format:

```
{
  "Narrative": ["attribute1:value1", "attribute1:value1", ...],
  "Visual Style": ["attribute1:value1", "attribute1:value1", ...],
  "Content Sources": ["attribute1:value1", "attribute1:value1", ...]
}
```

USER_CONTEXT (dynamic):

****Narrative**** attribute1:value1, attribute2:value2, ...

****Visual Style**** attribute1:value1, attribute2:value2, ...

****Content Sources**** attribute1:value1, attribute2:value2, ...

A.1.2 Outline Generation From Intent Tags. To generate an outline based on a user's existing tags on the slide deck steering canvas, we construct a call to GPT comprising the static `SYSTEM_PROMPT` and the dynamic `USER_CONTEXT` below. `USER_CONTEXT` is a string dynamically constructed from all active intent tags [attribute:value] for each tag group (Narrative, Visual Style, Content Sources). GPT is instructed to generate an outline in Markdown format, including the URLs of image reference tags if present.

SYSTEM_PROMPT (static):

You are an assistant to help users author slide presentations in PowerPoint.

You are creating an outline for the user's presentation.
 The user will have descriptions and keywords that describe different aspects of the presentation.
 The user will prompt you in the format: attribute1:value1, attribute2:value2, ...
 Ignore all the attributes related to the visual style.
 If the content sources include ImageUrl attributes, you should include these images in the outline as images in the markdown.
 Respond with an outline of the presentation in the format:
 section title - section content bullet points.
 Return the outline in markdown format without ```markdown at the start and end. Only return markdown content for the outline.

USER_CONTEXT (dynamic):

****Narrative**** attribute1:value1, attribute2:value2, ...
****Visual Style**** attribute1:value1, attribute2:value2, ...
****Content Sources**** attribute1:value1, attribute2:value2, ...

A.1.3 Slide generation mechanism. To generate the JSON structure for a slide deck, we construct a call to GPT comprising the static **SYSTEM_PROMPT** and the dynamic **USER_CONTEXT**. **USER_CONTEXT** is a string containing the **PRESENTATION OUTLINE** in markdown, **META INFORMATION**, which is a string containing all active intent tags [attribute:value], and optionally the **REFERENCE SLIDE DECK TEMPLATE**, which is a JSON string of a reference slide deck (if provided). The general slide deck template schema and slide generation mechanism are based on the Spectacle¹ library.

SYSTEM_PROMPT (static):

Your task is to generate a slide presentation deck from an outline and meta information.
 You will receive the presentation **OUTLINE** in markdown format from the user.
 You will also receive additional **META INFORMATION** about the desired presentation's Narrative, Visual Style, and Content Sources.
 The **META INFORMATION** will describe different aspects of the presentation.
 You will receive the **META INFORMATION** in the format:
****Bucket**** attribute1:value1 (optional description), attribute2:value2 (optional description), ...
 A bucket might be empty, in which case you should suggest concepts based on the other buckets.
 A value might not be associated with an attribute.
 The **META INFORMATION** might also contain **USER DOCUMENTS** with content that has to be included in the presentation.

GENERATING SLIDES

Your task is to generate a slide deck. Each slide has to consist of a **LAYOUT**, **CONTENT**, and **THEME**.

LAYOUT:
 The layout determines the arrangement of the content on the slide.
 There are five different layouts to choose from:
 "title", "listOrParagraph", "verticalImage", "fullImage".

CONTENT:
 Each layout takes different content parameters depending on the layout type. Optionally, each slide can take an image url as background image in the content.
 The background image will replace the background color of the slide.
 The "listOrParagraph" and "verticalImage" layouts take either a list or a paragraph as content besides the title and image.
 For bullet lists, never use more than 3 bullets.

THEME:
 The theme determines the fonts, colors, and font sizes used in the slide.

fonts: { header: fontname, text: fontname }
 "header" is the font used for titles and "text" is the font used for body, lists, paragraphs.
 There are six fonts to choose from: "Quicksand", "Playfair Display", "Montserrat", "Merriweather", "Roboto" and "Roboto Condensed".

colors: { primary: color, secondary: color, tertiary: color }
 "primary" is the main color used for text, "secondary" is the color used for accents, and "tertiary" is the color used for backgrounds.

fontSizes: { h1: size, text: size }

¹<https://github.com/FormidableLabs/spectacle>

"h1" is the font size used for titles and "text" is the font size used for body text.

REFERENCE SLIDE DECK TEMPLATE

The user might also provide a REFERENCE SLIDE DECK TEMPLATE with a slide deck template they want to use for the visual style. Important: If a REFERENCE SLIDE DECK TEMPLATE is provided, use only this template for the visual style (font, backgroundImage, sizes, colors) and IGNORE all visual style information from the generic example slide deck below. ALWAYS use the backgroundImage URLs from the REFERENCE SLIDE DECK TEMPLATE on each slide!

REWORK CONTENT FROM THE OUTLINE

You should rework the content from the OUTLINE to fit the slide deck based on the desired format specified in the META INFORMATION.

For example, if the META INFORMATION specifies that the presentation should have a formal tone, you should rework the content to fit this tone.

If the META INFORMATION specifies that the presentation should have a more verbose text, you should rework the content to fit this style.

YOUR TASK:

Your task is to compose a slide deck based on the outline and to adapt the content to the layout and theme.

Your task is to define the theme and layout of the slides based on the OUTLINE and META INFORMATION provided by the user.

Generate the number of slides based on the outline and meta information provided by the user.

Keep the visual style consistent across all slides.

Especially the fonts, colors, background color and font sizes.

You should return the slide deck in JSON format. Only return JSON format without any additional Markdown wrapper code blocks.

Here is a generic example of a slide deck in JSON format:

```
[{
  slideNumber: 1,
  layout: "title",
  content: {
    title: "A presentation about something",
    subtitle: "by someone",
    backgroundImage: "url(<image url placeholder>)"
  },
  theme: {
    fonts: {
      header: "'Playfair Display', serif",
      text: "'Quicksand', sans-serif",
    },
    colors: {
      primary: "#000",
      secondary: "#000",
      tertiary: "#fff",
    },
    fontSizes: {
      h1: "100px",
      text: "44px",
    },
    space: [16, 24, 32],
  },
}, {
  slideNumber: 2,
  layout: "listOrParagraph",
  ...
  (PARTS OMITTED)
  ...
}]
```

USER_CONTEXT (dynamic):

PRESENTATION OUTLINE:

\$CURRENT_OUTLINE_IN_MARKDOWN_STRING

```

META INFORMATION:
$CURRENT_ACTIVE_TAGS_AS_ATTRIBUTE:VALUE_PAIRS_CLUSTERED_BY_GROUP

(optional) REFERENCE SLIDE DECK TEMPLATE:
$REFERENCE_SLIDE_DECK_JSON

```

A.1.4 Mechanism to create tags for an existing slide (tag grounding act). To generate a collection of intent tags based on a provided image of the current slide, we construct a call to GPT comprising the static SYSTEM_PROMPT and the dynamic USER_CONTEXT below. USER_CONTEXT is a string of the current slide in a base64 image format. In addition to the slide image, we experimented with including the active steering board tags in the prompt but refrained from including these in this version of IntentTagger for simplicity. Future explorations outside the scope of this paper should investigate how "global" steering board-level intent tags can meaningfully propagate to individual slides and how changes on single slides can propagate back to the global steering board level while meaningfully resolving conflicts between global and slide-level tags.

```

SYSTEM_PROMPT (static):
You are an assistant to help users author slide presentations in PowerPoint.
You will receive an image of a slide from their slide presentation deck.
Your task is to analyze the slide deck and return descriptive attributes that best describe the slide.
You should return these attributes related to three categories: Narrative, Visual Style, and Content Sources.
Each attribute should consist of a label and a value, like Tonality:Formal or Typography:Modern.
For each bucket, return between 2 to 6 attributes.
Return your response in JSON format.
Only return JSON format without any additional Markdown wrapper code blocks.

Here is an example of the JSON format:
{
  "Narrative": ["attribue1:value1", "attribue1:value1", ...],
  "Visual Style": ["attribue1:value1", "attribue1:value1", ...],
  "Content Sources": ["attribue1:value1", "attribue1:value1", ...]
}

USER_CONTEXT (dynamic):
$IMAGE_OF_CURRENT_SLIDE_AS_BASE64_STRING

```

A.2 Semi-structured Task Outcomes

This section contains screenshots of IntentTagger (deck steering board and slide panel) taken at the end of each participant's semi-structured slide creation task from our user study (see study phase 3 at section 5.2).

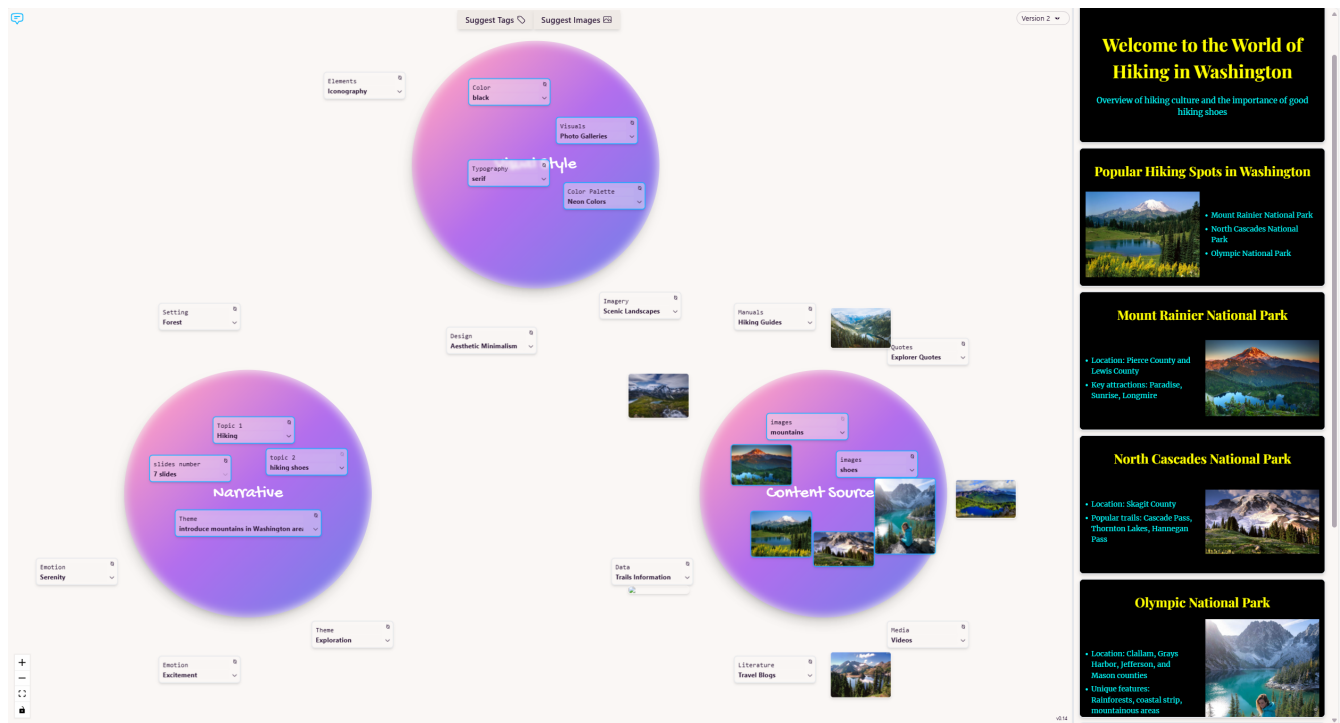


Figure 19: Semi-structured task outcome of P01

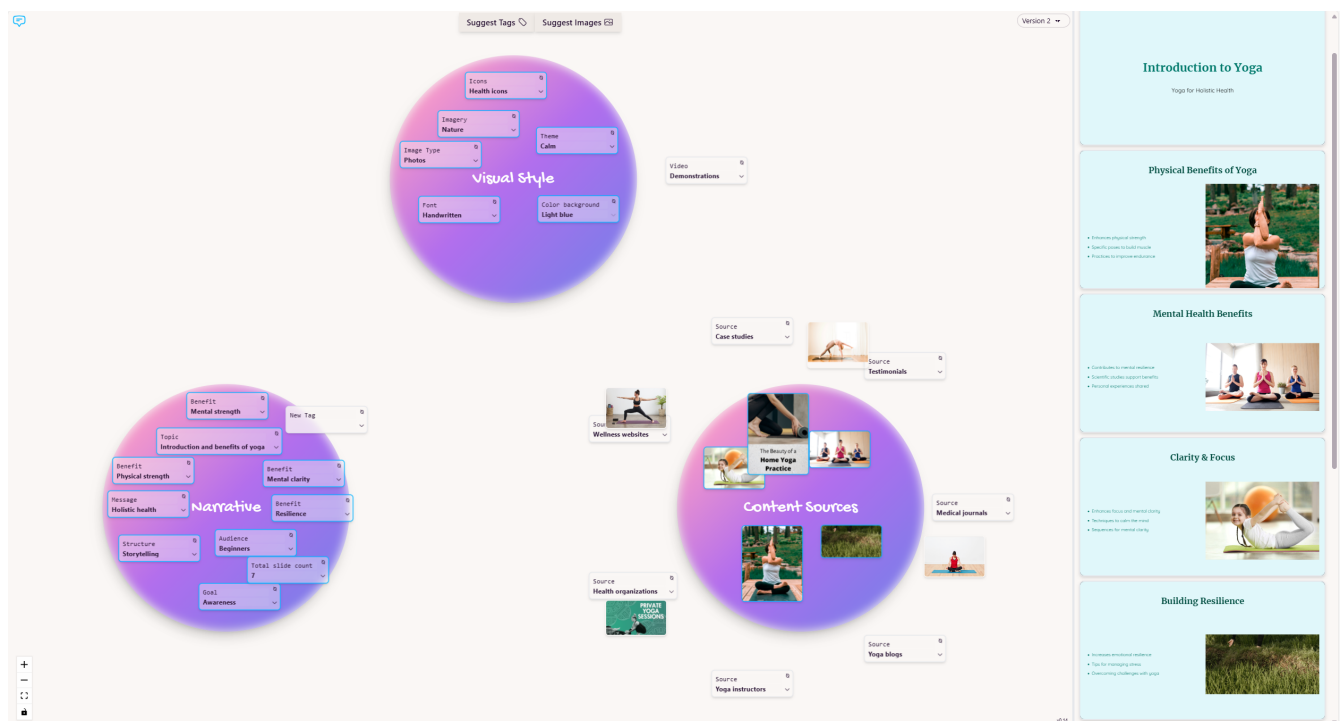


Figure 20: Semi-structured task outcome of P02

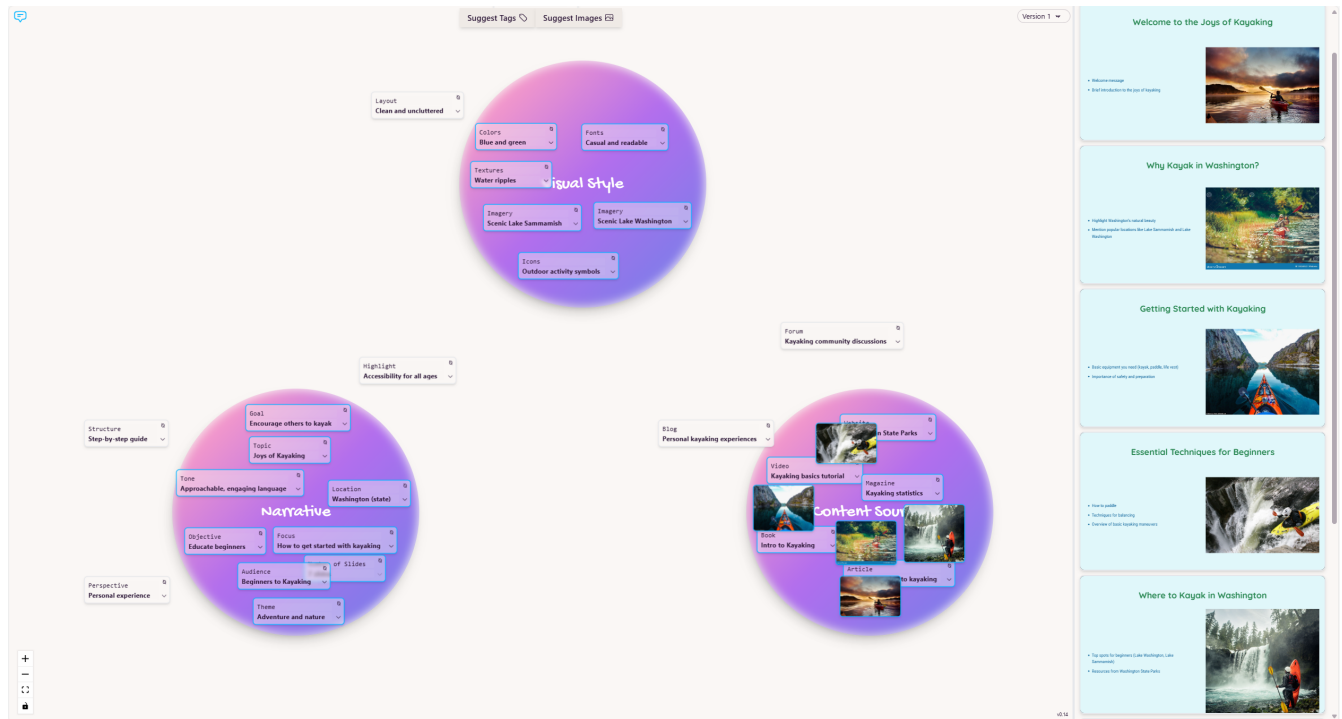


Figure 21: Semi-structured task outcome of P03

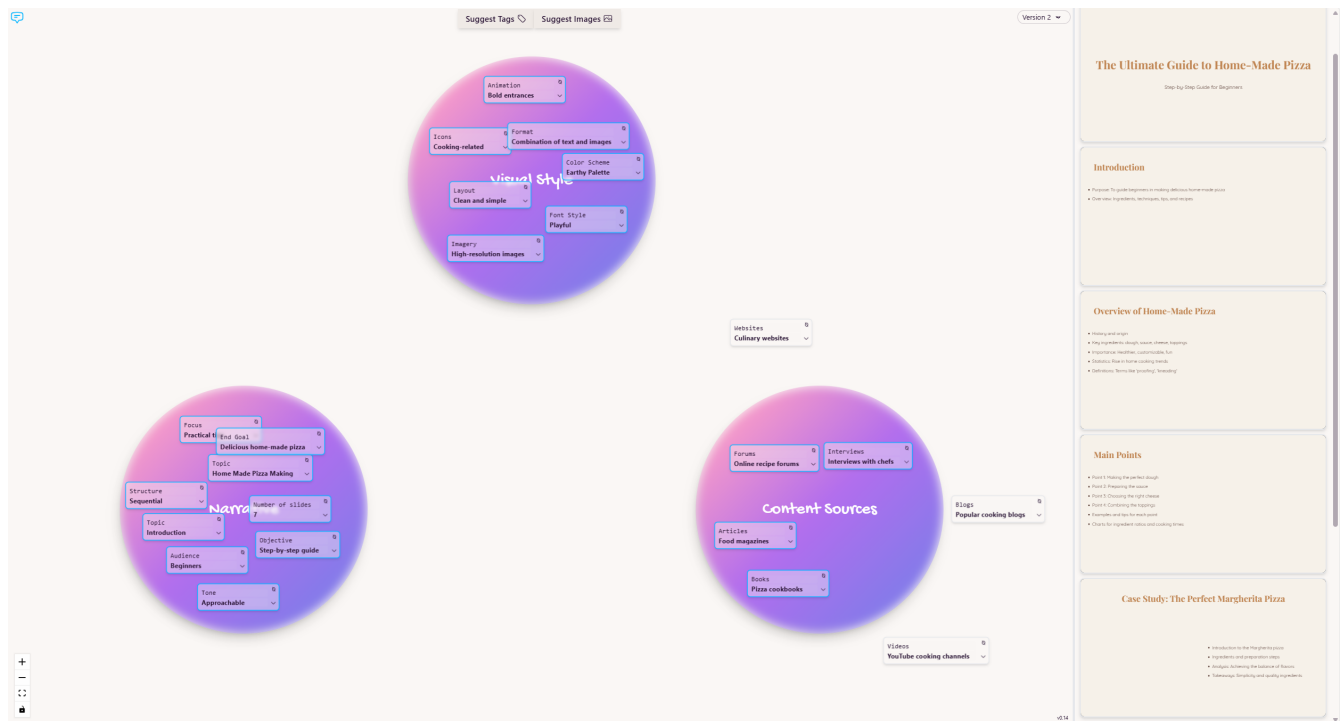


Figure 22: Semi-structured task outcome of P04 (Note: The user was unable to add images to their presentation, due to a temporary outage of the image search API.)

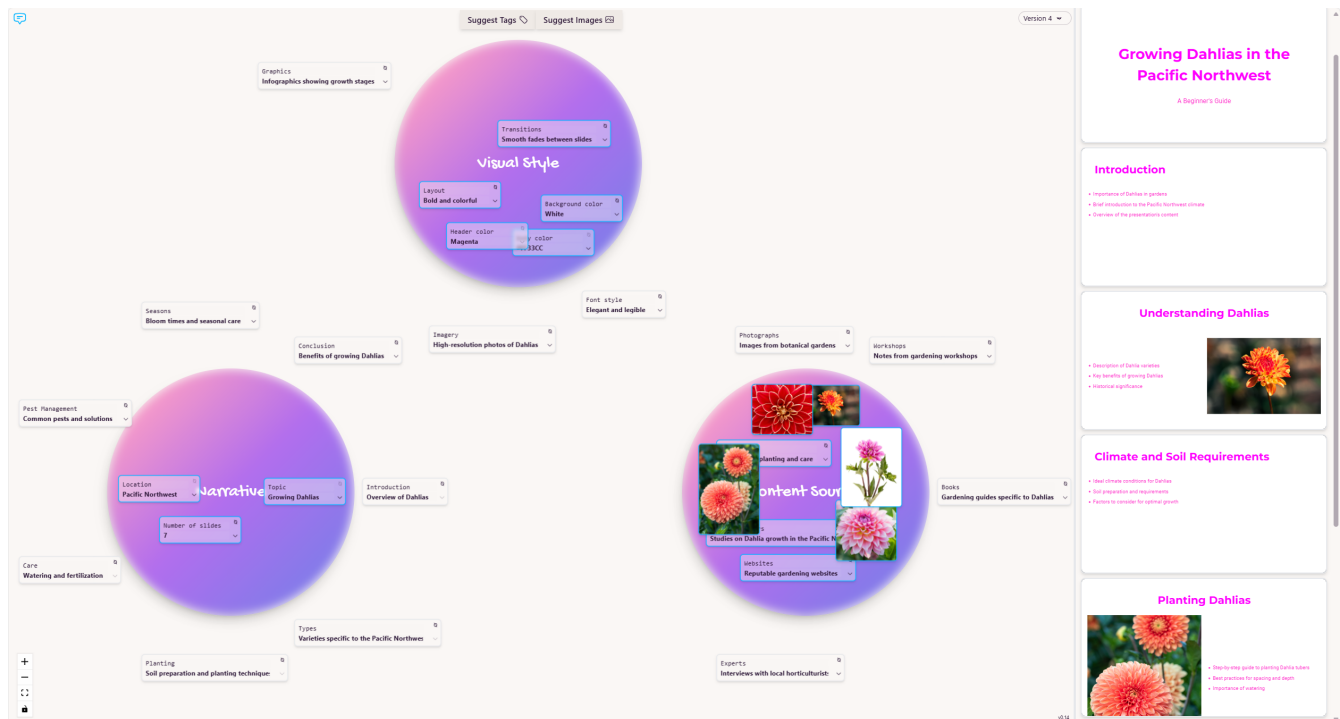


Figure 23: Semi-structured task outcome of P05

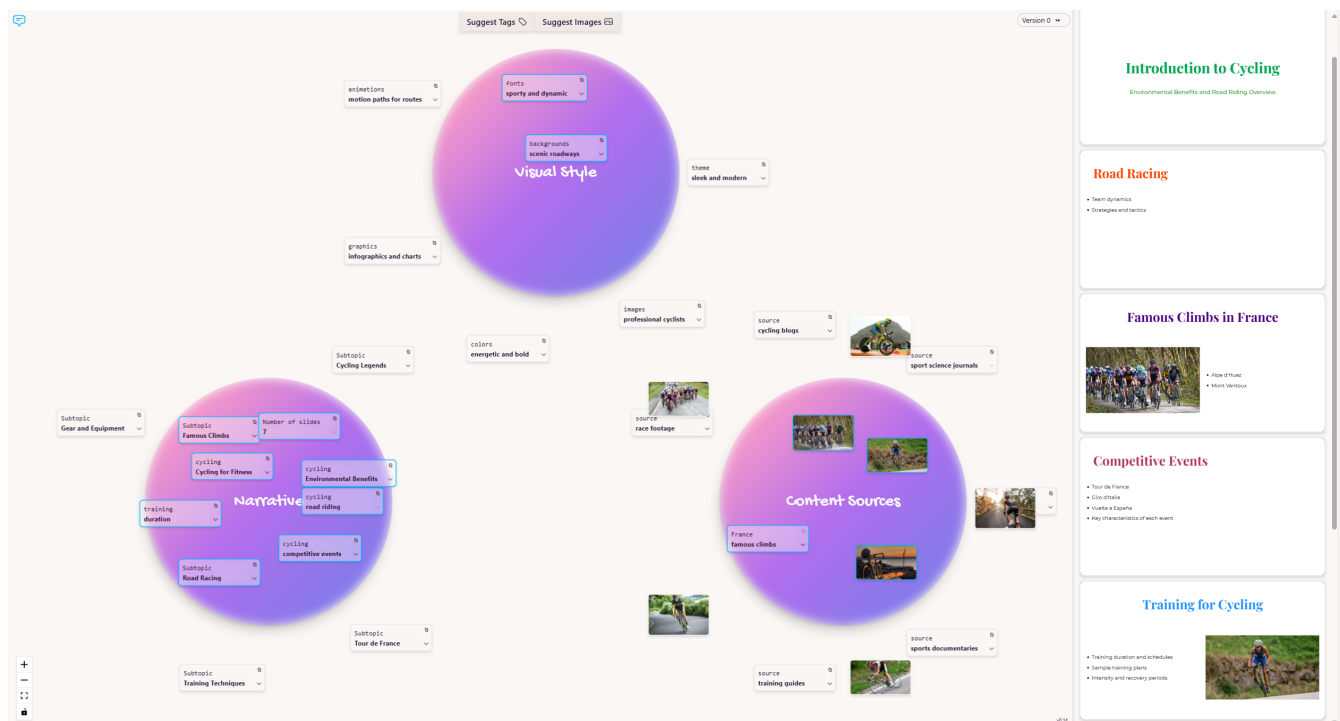


Figure 24: Semi-structured task outcome of P06

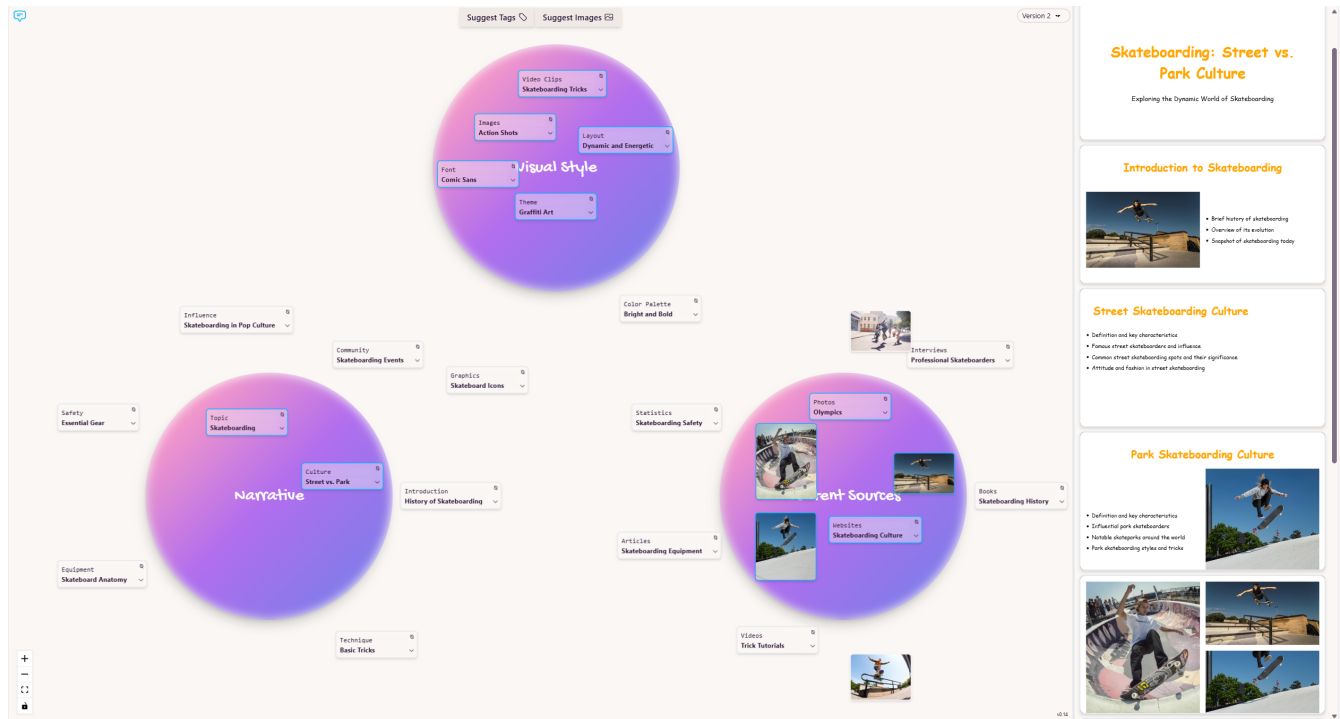


Figure 25: Semi-structured task outcome of P07

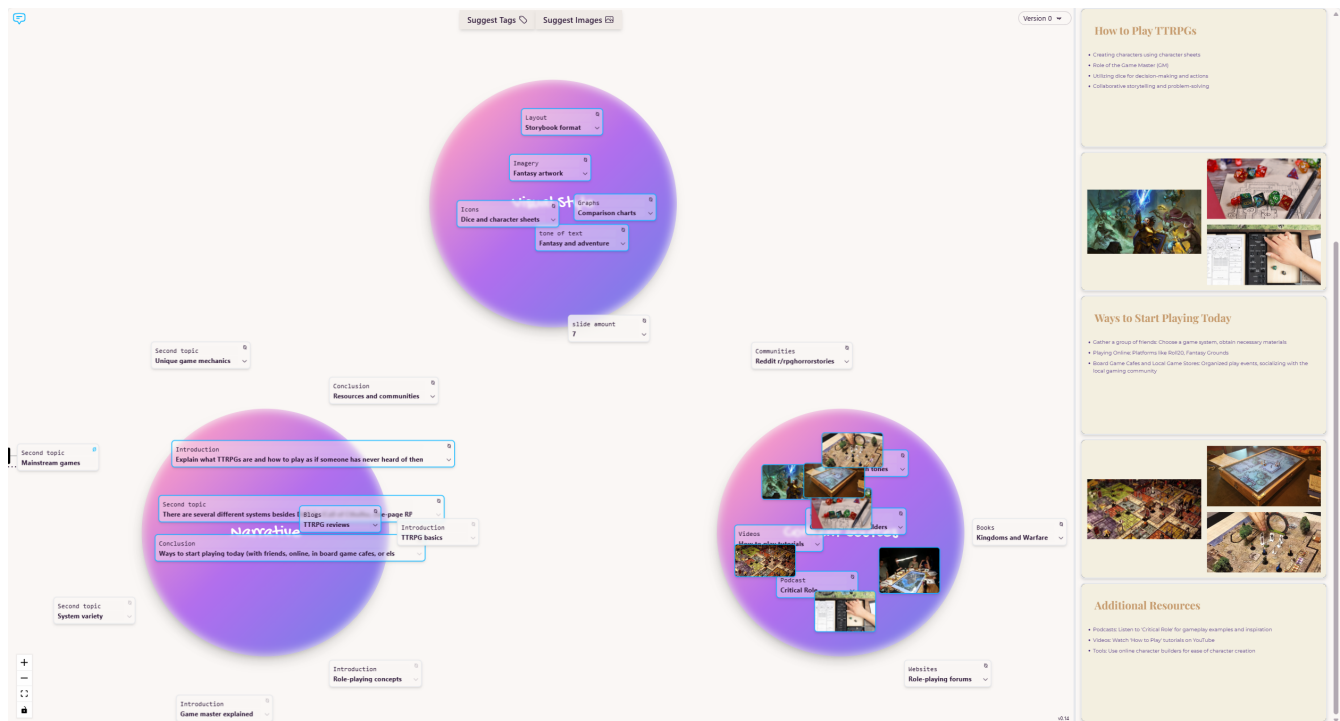


Figure 26: Semi-structured task outcome of P08

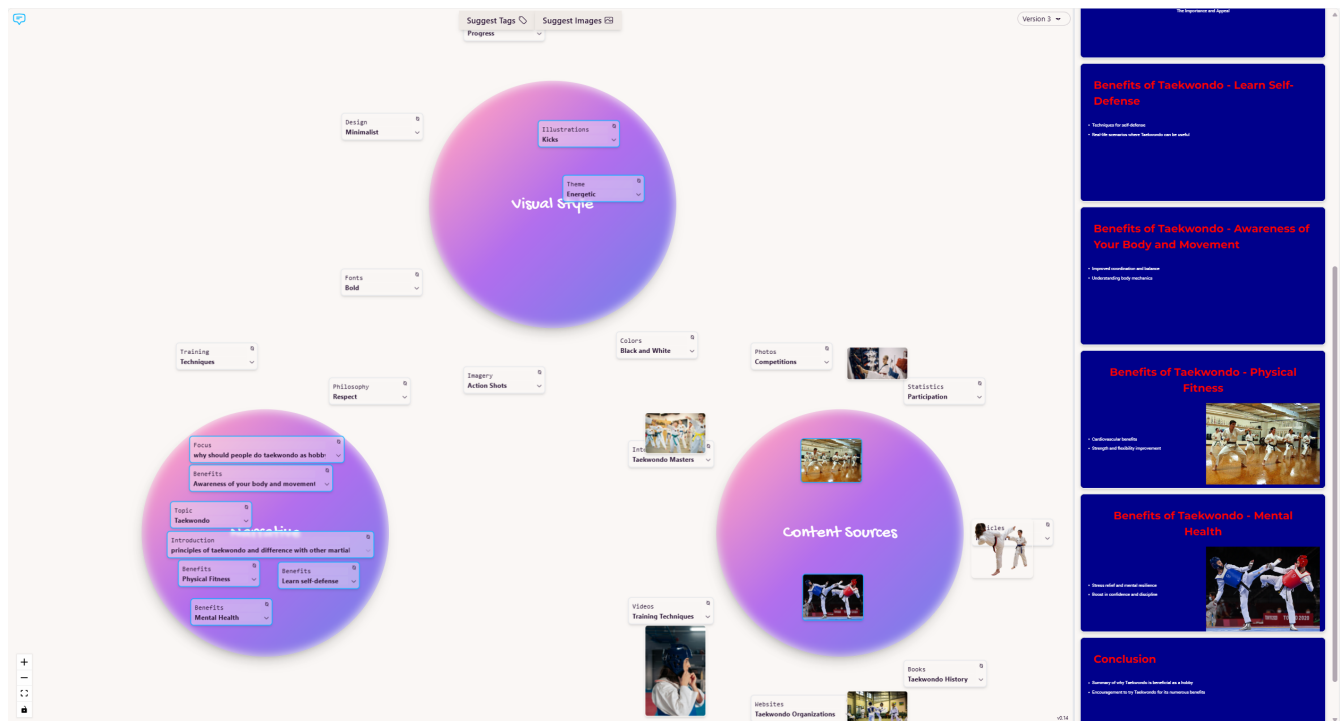


Figure 27: Semi-structured task outcome of P09

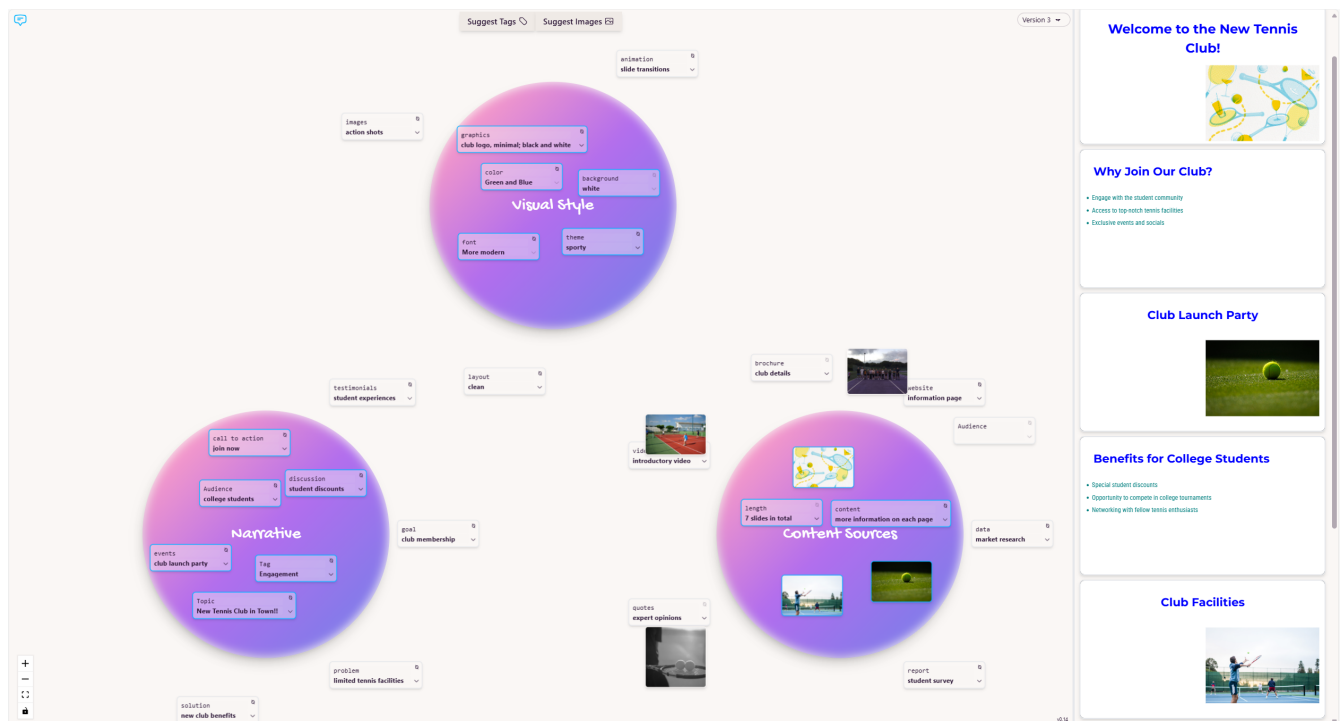


Figure 28: Semi-structured task outcome of P10

