

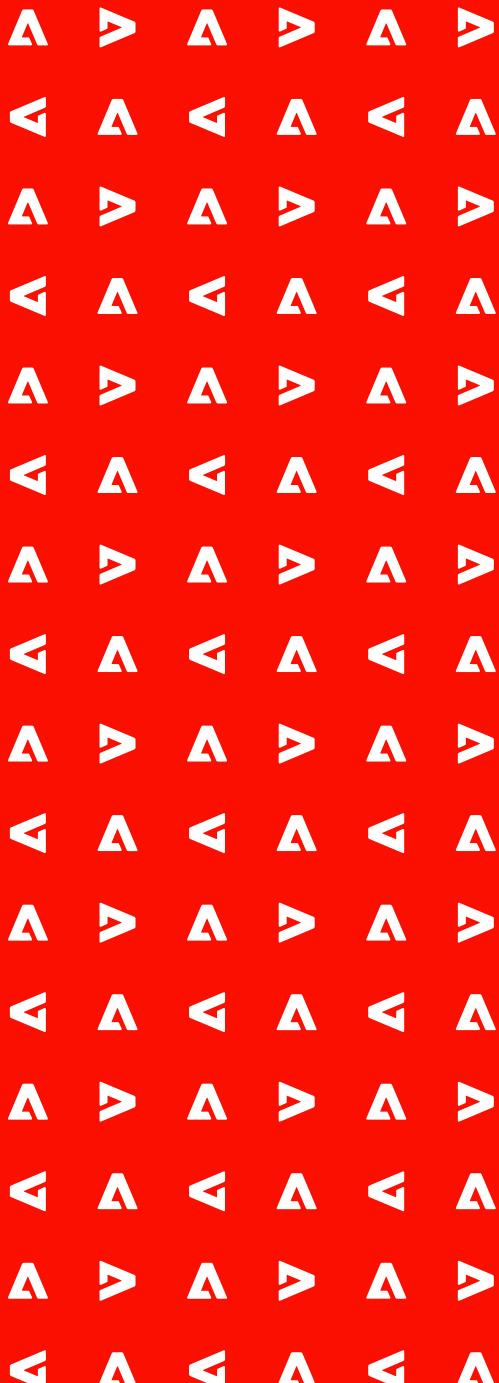


# Powering Creative Asset Recommendations with Hybrid Multi-Modal Search

Tracy Holloway King

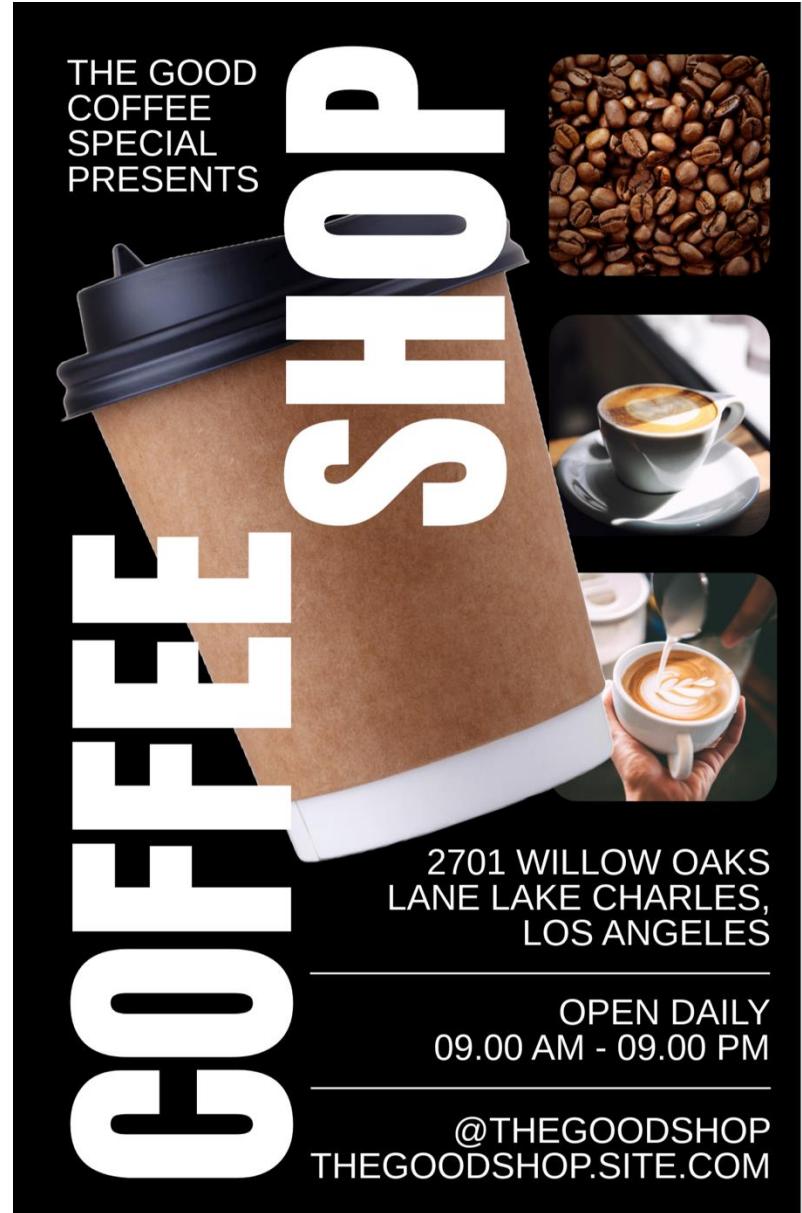
Adobe

2024-10-25



# Intro: Adobe Express Multi-modal Creation

- Creating marketing & social media content is challenging
- Content is multi-modal (multiple images & text)
- Recommendations based on the project:
  - Spark new ideas
  - Discover relevant content
  - Complete projects more quickly
- Contextual recommendation system:
  - Symbolic & embedding intent understanding
  - Domain-specific knowledge graph
  - Understand the creator's project
  - Determine recommended categories
  - Leverage multi-modal search
  - Provide asset recommendations that complement the project



# Adobe Express Project Creation: Start from Template

The screenshot displays the Adobe Express web application interface. At the top, there's a search bar with placeholder text "Search for content or generate your own". Below the search bar are three template preview cards: "Letter size (8.5 x 11 inches)", "Premium member", and "Join community". On the left side, a vertical sidebar menu includes icons for Home, Your stuff, Brands, Templates, Schedule, Learn, and Add-ons. The main content area is titled "Browse templates" and shows a grid of 12 different project templates across four rows. The templates include:

- Employee Satisfaction Survey 20XX**: A green-themed survey report with statistics like 83% and 79%.
- FITNESS FASHION RACE IN STYLE**: An advertisement for running gear with a blue and yellow color scheme.
- SPORTING OUTFITTERS**: An advertisement for hiking essentials featuring a yellow and black backpack.
- GAME STORE LEVEL UP YOUR GAMING**: A vibrant purple-themed gaming advertisement.
- Find HR Support Today**: A small card for HR services.
- FINANCIAL YEAR 20XX CONFERENCE**: A dark-themed conference announcement.
- RH Graduate Scheme Testimony**: A testimonial from Yassir Denton about his experience.
- GRAC Hardware SHOP SMART, SAVE BIG**: An advertisement for hardware tools.
- Commercial Real Estate Report 20XX**: A report cover for commercial real estate.
- Revamp Your Toolbox With Our Essentials**: An advertisement for tool essentials.
- JUST CLOSED**: A real estate closing announcement for a house in Beacon, NY.

At the bottom of the page, there are "Recent" project thumbnails and a "View all" link.

# Adobe Express Project Creation: Start from Text

The screenshot shows the Adobe Express web interface. On the left, there's a sidebar with various sections like Home, Your stuff, Brands, Templates, Schedule, Learn, and Add-ons. A prominent 'Browse templates' section displays a 'Employee Satisfaction Survey 20XX' template for 'Local Fashion'. The main workspace is titled 'Untitled - October 22, 2024 at 13:52:06' and shows a document page with the text 'Page 1 / 1 - Add title'. Below the text, there's a large, bold heading 'Coffee Shop Soft Opening'. To the left of the workspace, a 'Media' panel is open, showing sections for Photos, Video, and Audio. The Photos section contains two thumbnail images: one of a person in a yellow shirt and another of a person working on a computer. The Video section shows a thumbnail of a campfire. The Audio section lists three audio files: 'Healthy Business - Loop', 'Gentle Daydreams - Loop A', and 'Silent Solo Night loop'. The top navigation bar includes a search bar, community links, and account information for a 'Premium member'.

# Adobe Express Project Creation: Start from Photo

The screenshot displays the Adobe Express interface, showing two projects in progress.

**Left Project:** Untitled - October 22, 2024 at 13:52:06. This project is based on a template for an "Employee Satisfaction Survey 20XX". It features a green background with white text and several large, bold statistics in ovals. The stats include "83%" (I am proud to work at Local Fashion), "83%" (I can see myself working at Local Fashion in 2+ years' time), "79%" (Most of the systems and processes here support us getting our work done effectively), and "75%" (I receive appropriate recognition when I do good work). A small image of two people in hard hats is visible at the bottom. The Adobe Express sidebar on the left shows this project under "Recent".

**Right Project:** Untitled - October 21, 2024 at 08:19:17. This project is titled "Page 1 / 1 - Add title". It features a large image of a woman pouring coffee into a cup at a cafe counter. The Adobe Express sidebar on the right shows this project under "Recent".

The top navigation bar includes the Adobe Express logo, a search bar, "Join community", "Premium member" status, and user profile icons. The toolbar at the top of each project window includes "Download" and "Share" buttons.

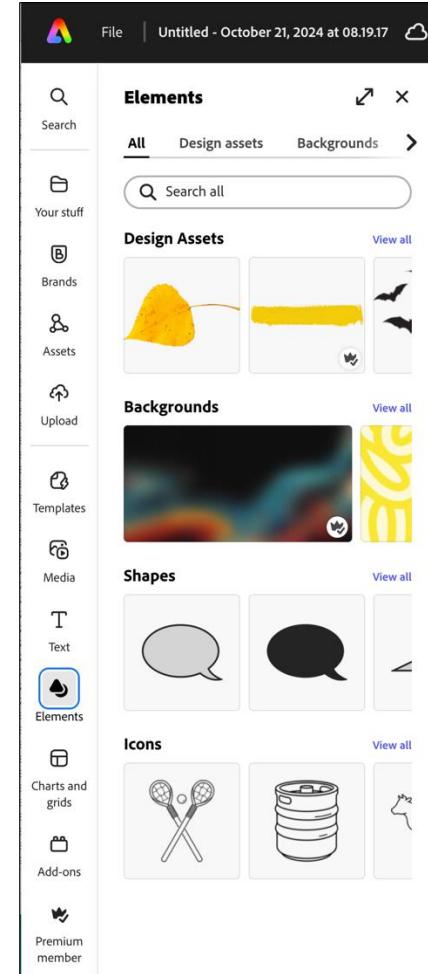
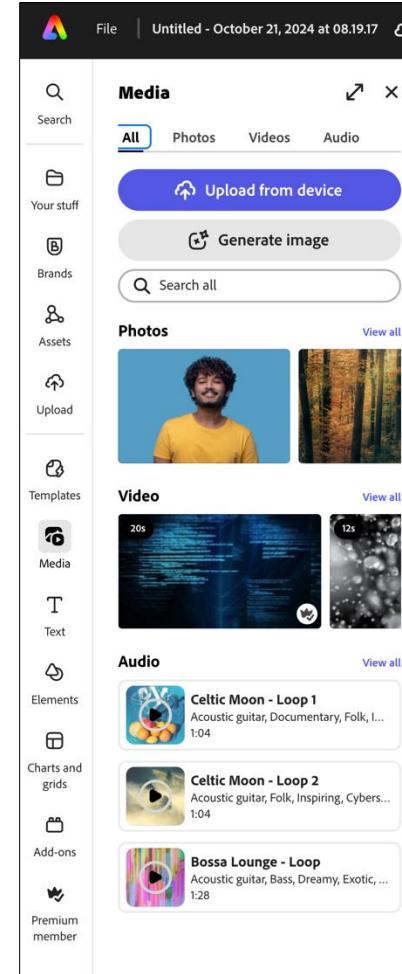
# Recommendations of Assets

## ▪ Current situation

- Static left rail categories
- Media and Elements separate

## ▪ Challenges

- Text and image canvas
- Not professional designers
- Latency and privacy
- Mobile screen real estate



- ## ▪ Goal: Provide asset recommendations based on the user's canvas to improve time to completion.

# Components Needed

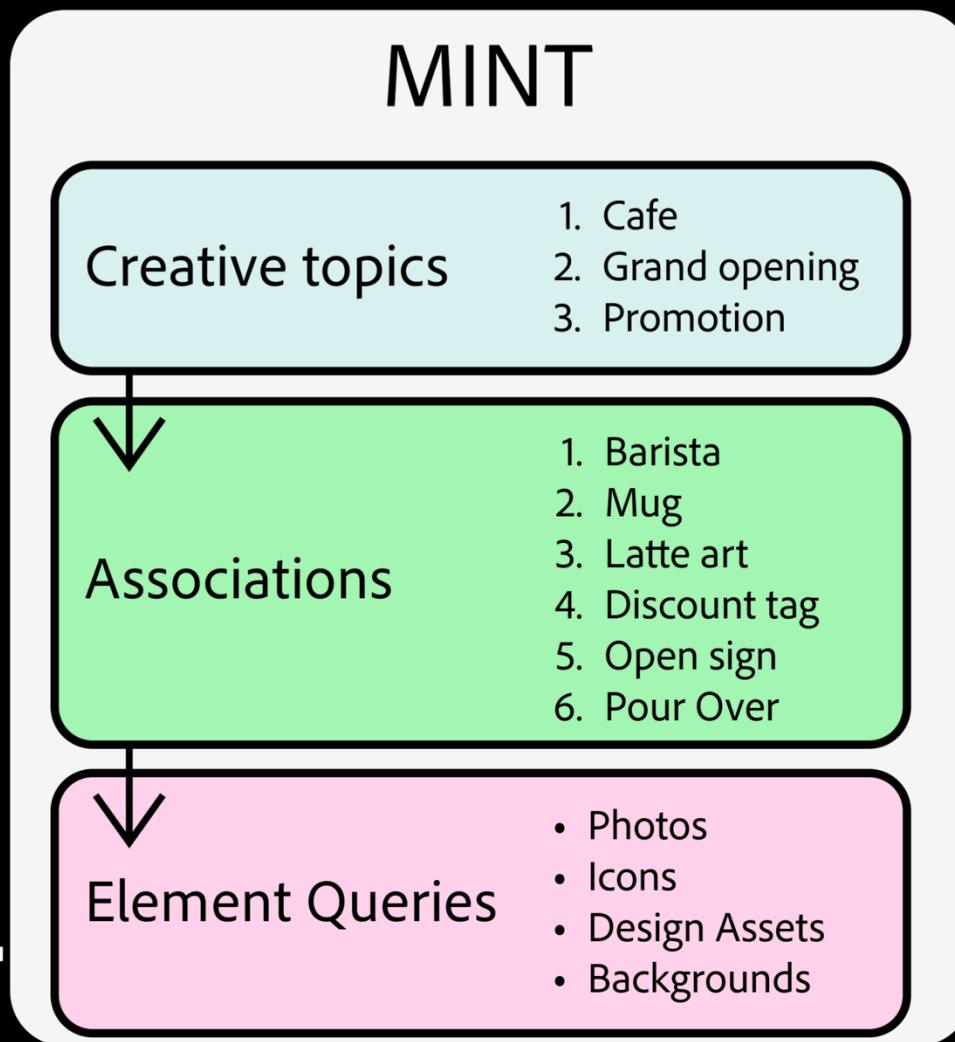
- Canvas understanding
  - Combined image and text
  - Low latency
- Mapping from canvas understanding to:
  - Asset types (photos, icons, etc.)
  - Intents with interpretable labels
  - Recommended assets
- UX enablement

The screenshot shows the Adobe Elements interface with the 'Icons' tab selected. A search bar at the top says 'Search icons'. Below it, a section titled 'Ideas based on your file' displays three recommended icons: 'Coffee' (a steaming cup), 'Cash register' (a cash register machine), and 'Tea' (a teapot and cup). Further down, another 'Ideas based on your file' section shows 'Tea' (a teapot and cup) and 'Restaurant' (a logo with 'WO' and 'Restaurant'). A blue circle highlights the 'Restaurant' icon. The left sidebar lists categories: Search, Your stuff, Brands, Assets, Upload, Templates, Media, Text, and Elements. The right sidebar lists categories: Coffee, Cash register, Tea, Restaurant, and Mug, each with a 'View all' link.

1. Canvas rendition and inner text  
are sent to MINT



2. MINT infers the intent of the canvas

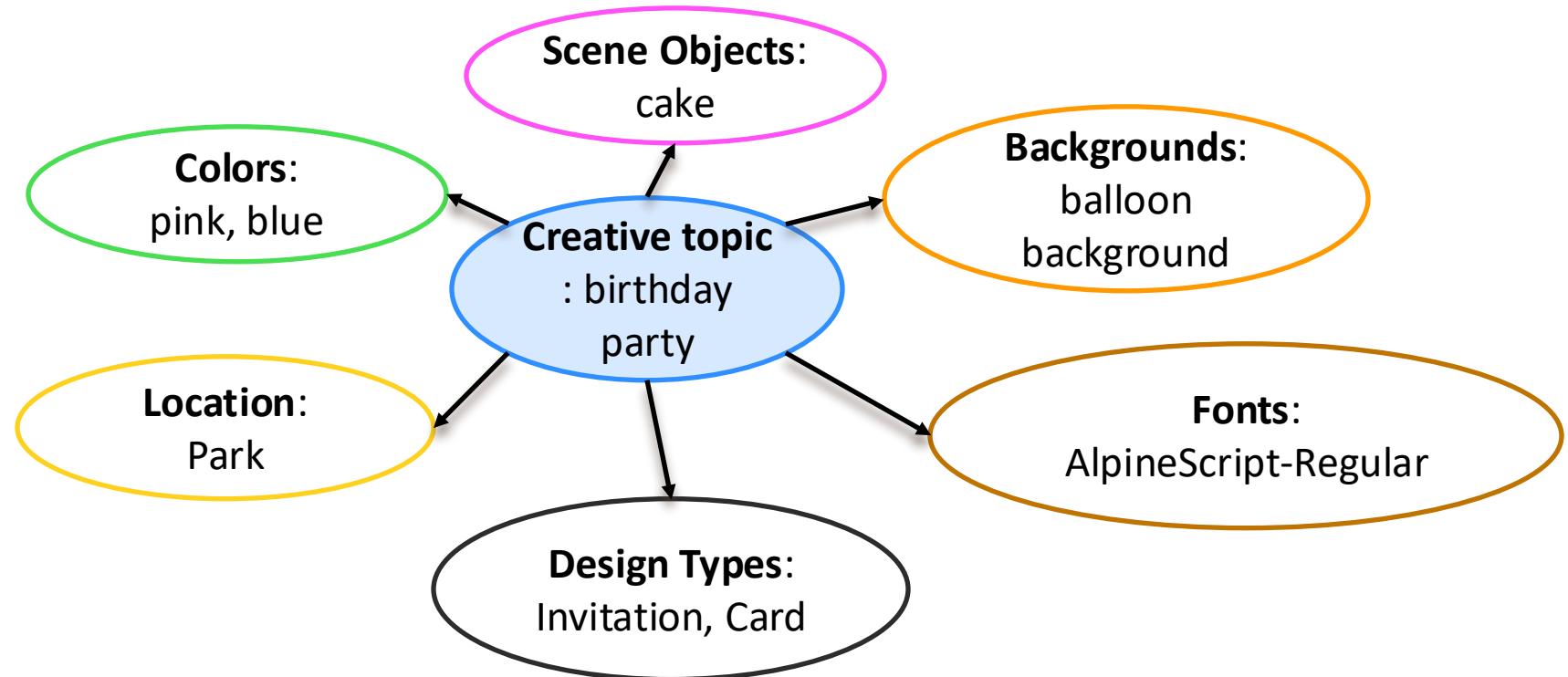


4. Asset queries are sent to USS

3. MINT looks for the  
related associations  
around these topics

# Creative Knowledge Graph

- CKG connects the user's intent to key aspects of creative process including:
  - Foreground and background image classes, font families, colors, icons
- ~120K intent nodes (typed)
- ~1.2M edges (scored & typed)
- MINT service
- Graph API service

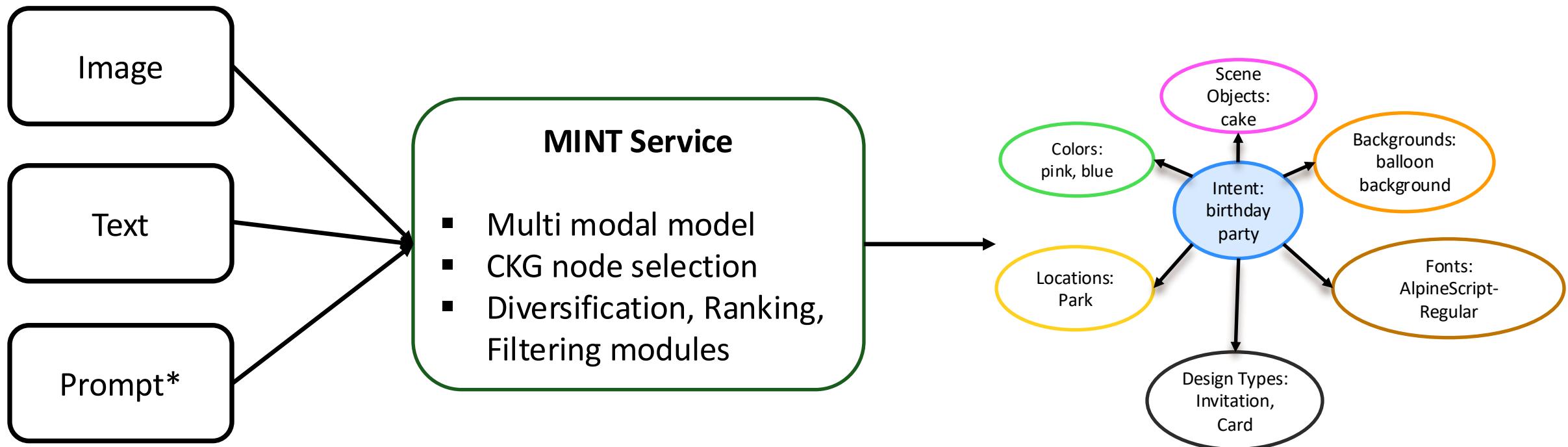


# Creative Knowledge Graph

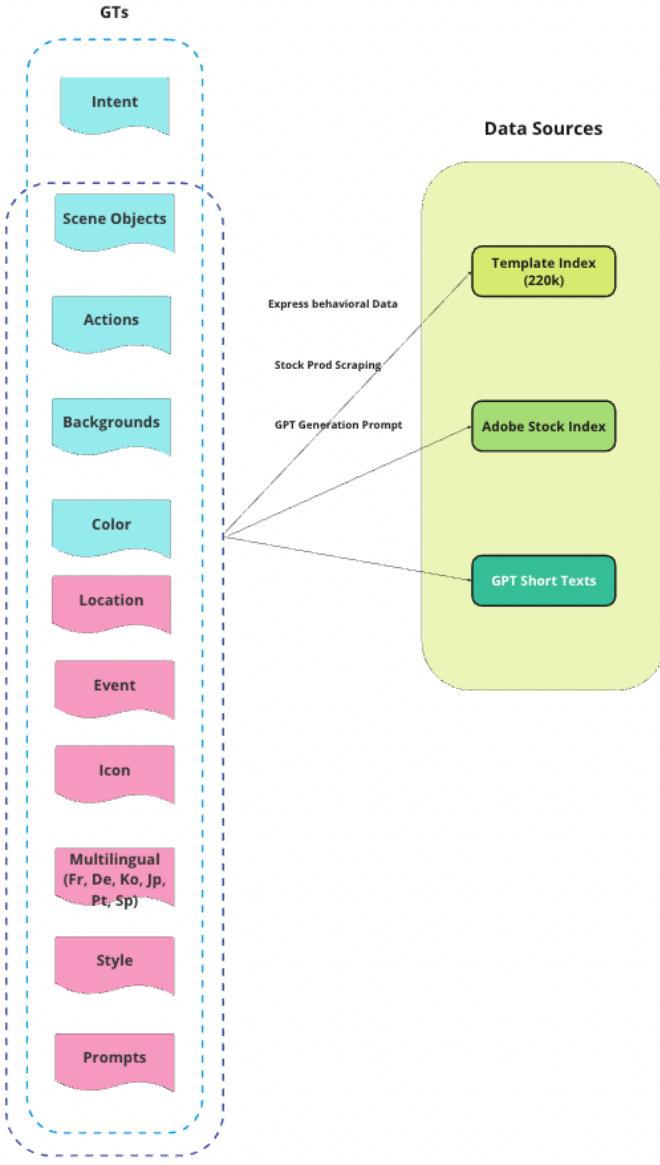
- A label-aligned multi-modal model that can:
  - Map to discrete (symbolic) nodes
  - Provide a semantic embedding
  - Handle all types in CKG (topic, background, colors, icons, etc.)
- Model architecture
  - A label-aware supervised contrastive training framework for multi-modal intent detection
  - A label-aware loss function (**SupCoLA**) for greater alignment to labels for detection, classification and embedding matching tasks
  - A scalable approach to create a large training sets for supervised learning

# Intent Understanding

- Input: multiple modalities from project (or user query)
- Output: Relevant graph nodes with scores



# Datasets for Model Training



Global Types	Dataset Size	Dataset Sources
Creative Intent	11.4M	Stock Asset (title, image) Template Asset (template metadata, rendition) GPT Texts
SceneObject, Background, Icon, Color, Scene Action, Location, Event	12.8M	Stock Asset (title, image) GPT Texts
Creative Intent, Background, Scene Object, Action, Color	250K	Firefly Prompts + Text2Template Prompts + Stock Prompts <i>* only texts</i>
Background, Scene Object, Icon	3M	Template CML Documents (BLIP Captions + Stock Asset image)
Image Styles	1.1M	Stock Asset (title, image) GPT Texts
Creative Intent, Scene Object	22M	Multilingual Dataset (fr, de, sp, ko, jp, pt) <i>* translated GPT Texts + Stock Images</i>
TOTAL	~50M	

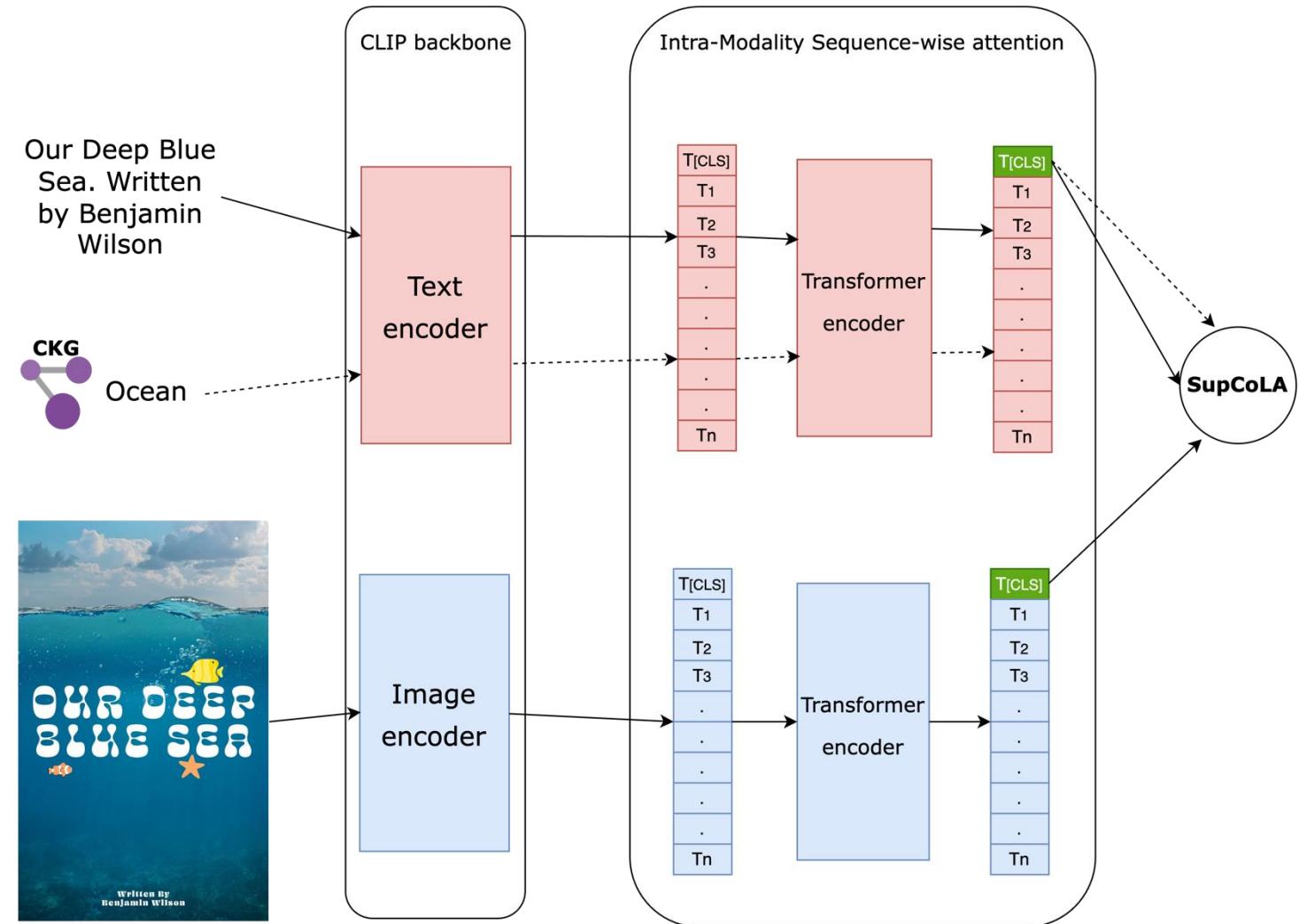
# Multi-Modal Model Architecture

- **CLIP backbone**

- openai/clip-vit-large-patch14-336
- Frozen weights. Last hidden state.
- Text [77\*768], Image [577\*1024]

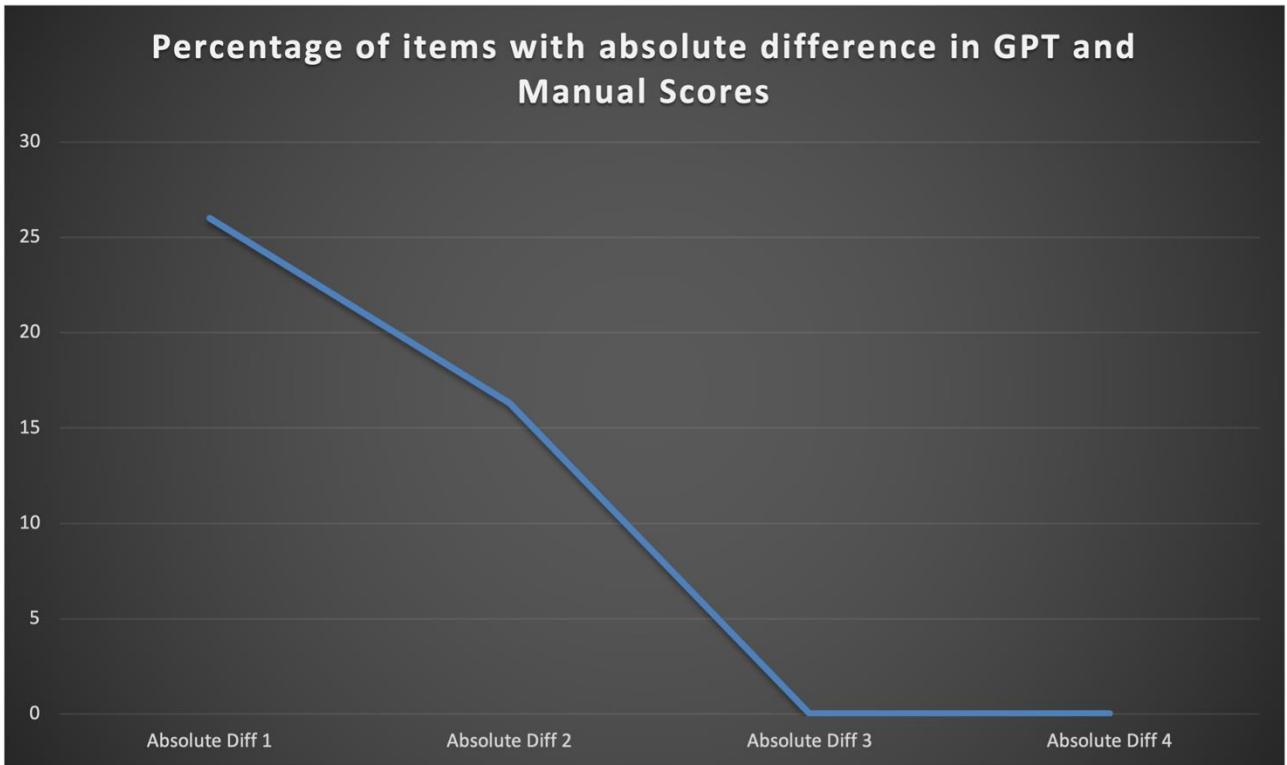
- **Sequence-wise self-attention blocks inspired from CMA-CLIP architecture**

- Lean multi-head transformer encoders on top of CLIP backbone.
- Applies self-attention to individual modality
- Finetune image or text encoders independently based on use cases.

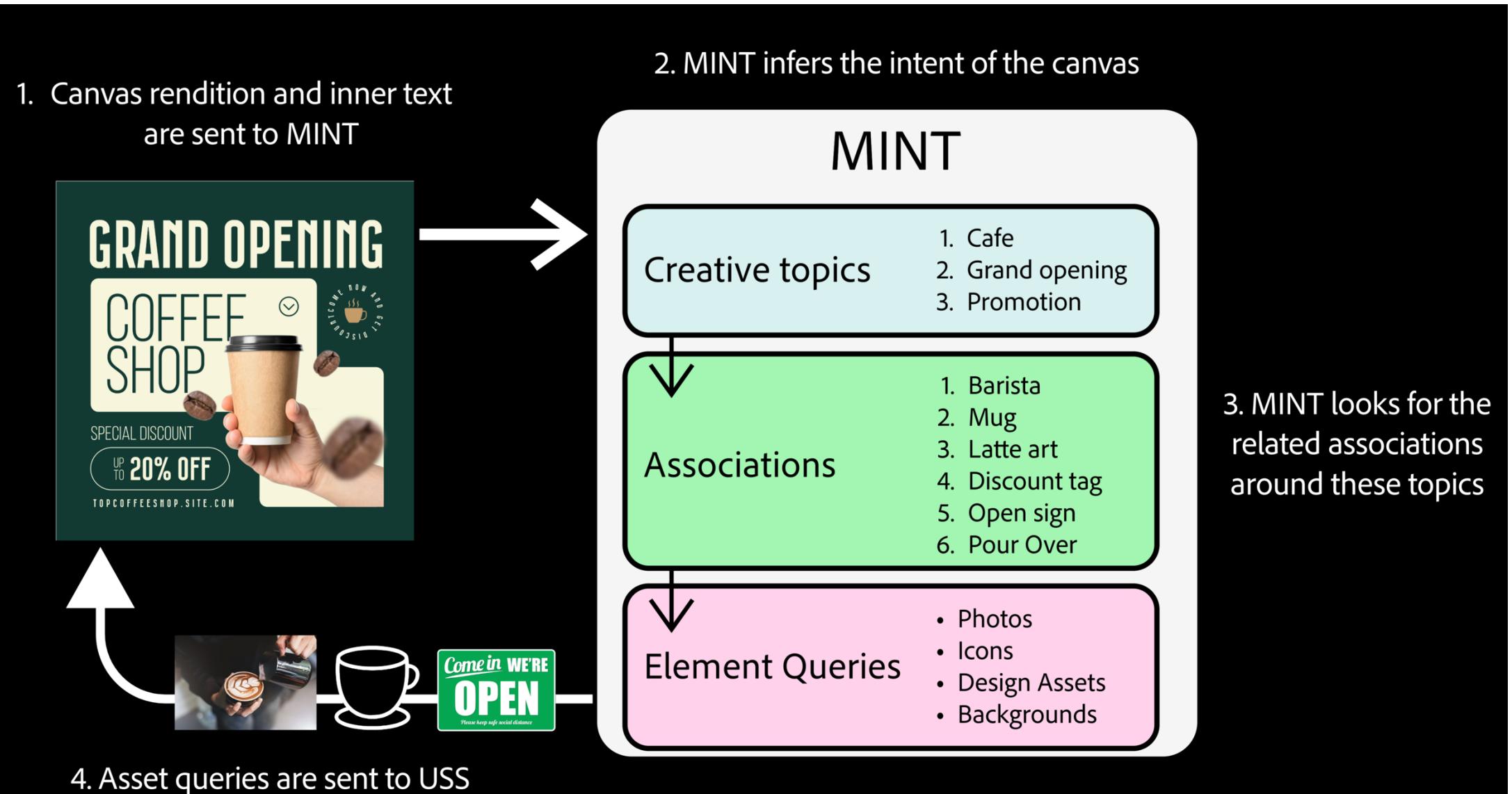


# CKG Semi-Automatic Evaluation

- GPT-4 and GPT-4V
- LLM evaluators benefits:
  - Cost & time: Multiple large evaluations
  - Consistency across runs
- Template-to-Intent evaluation
  - Relevance: scale 1-5
  - Diversity: scale 1-3
  - High correlation between human & LLM
    - 4380 judgements from 12 human evaluators
    - Pearson Score: 0.38



# CKG for Contextual Recommendations



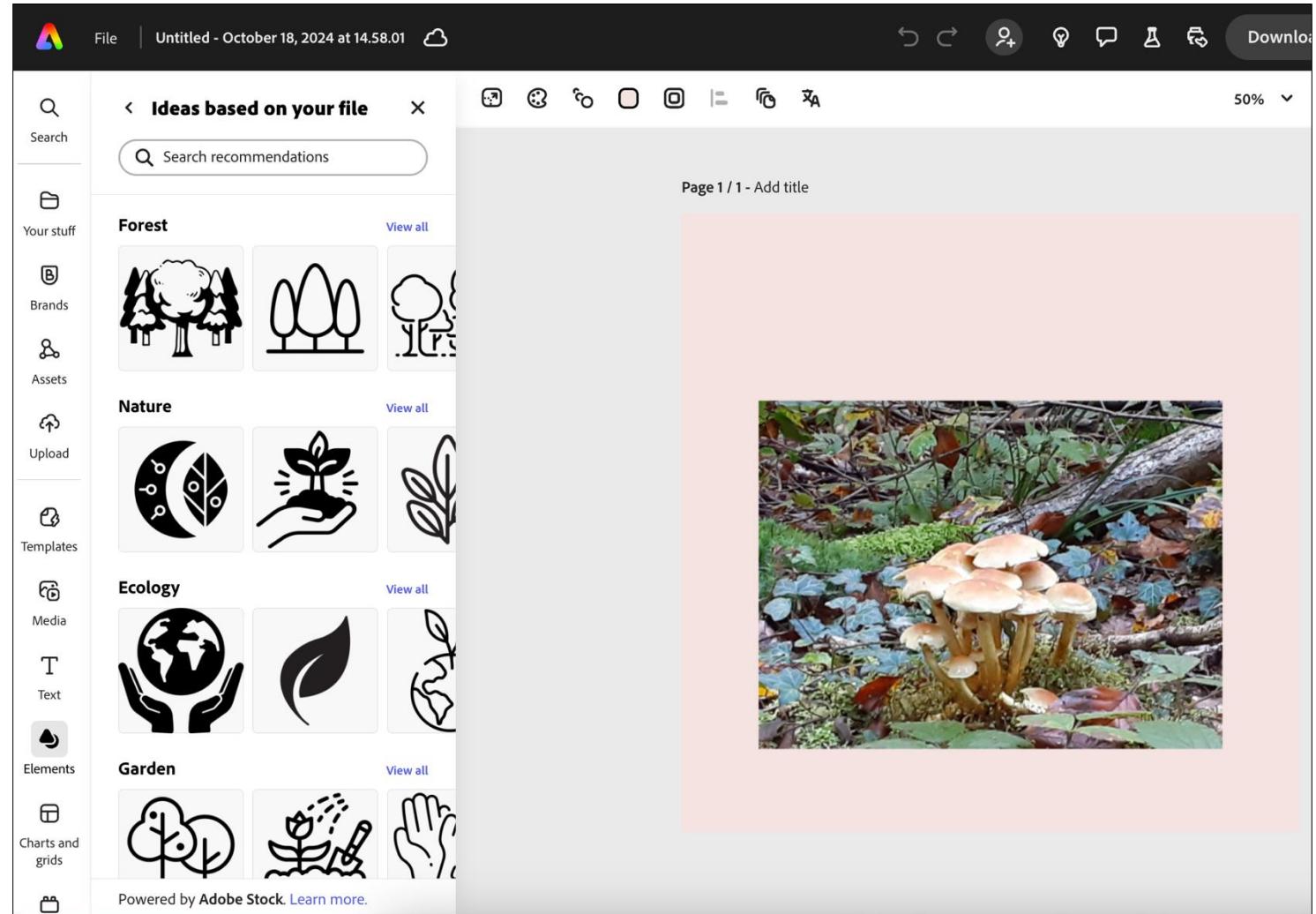
# Asset Recommendations

- One carousel of recommendations per CKG intent

- Displayed to user

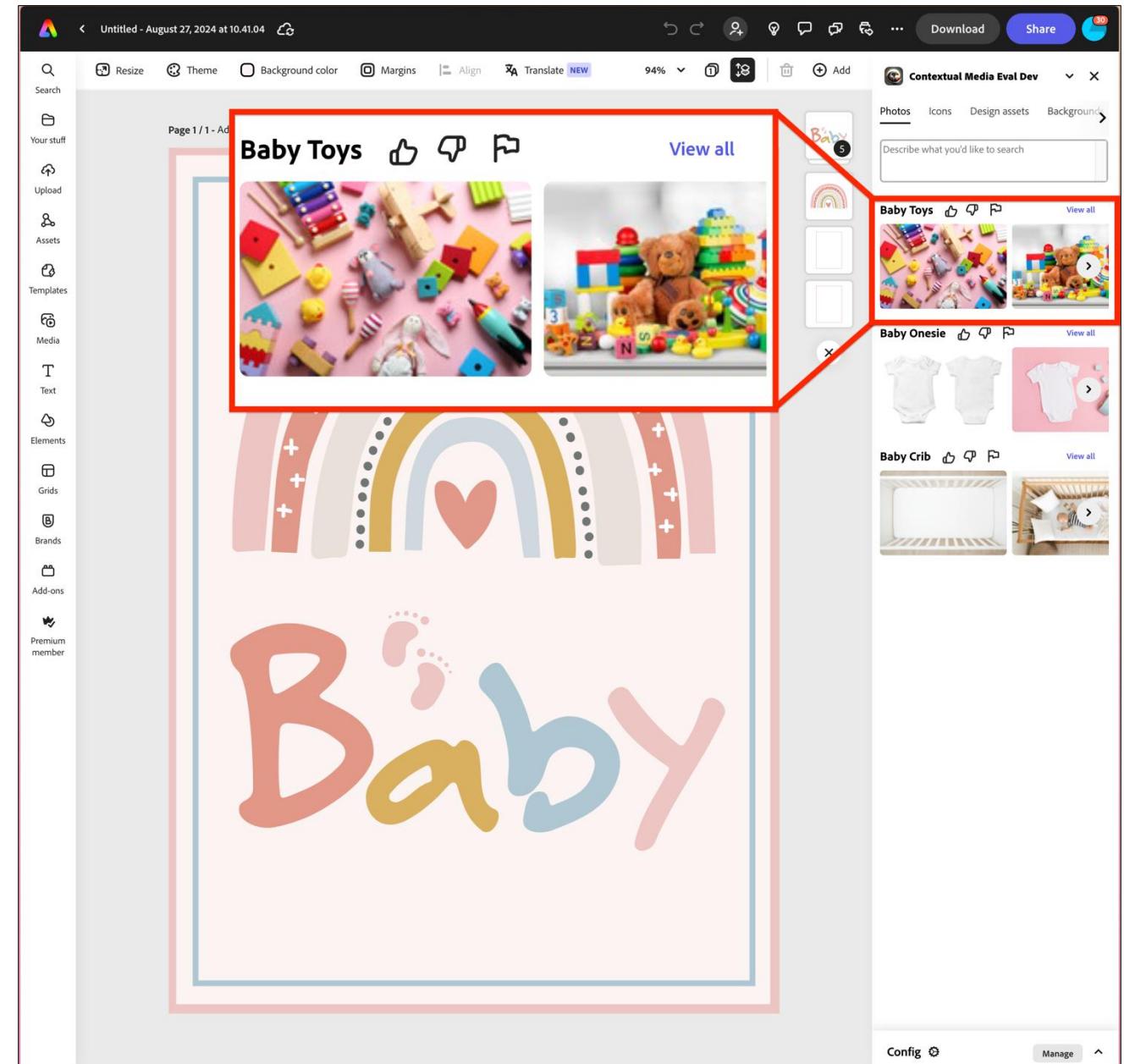
- Recommended assets via search

- CKG intent as text search query
  - Asset type (icons, backgrounds, etc.) constrained by UX search filters
  - Only top 2-3 assets visible



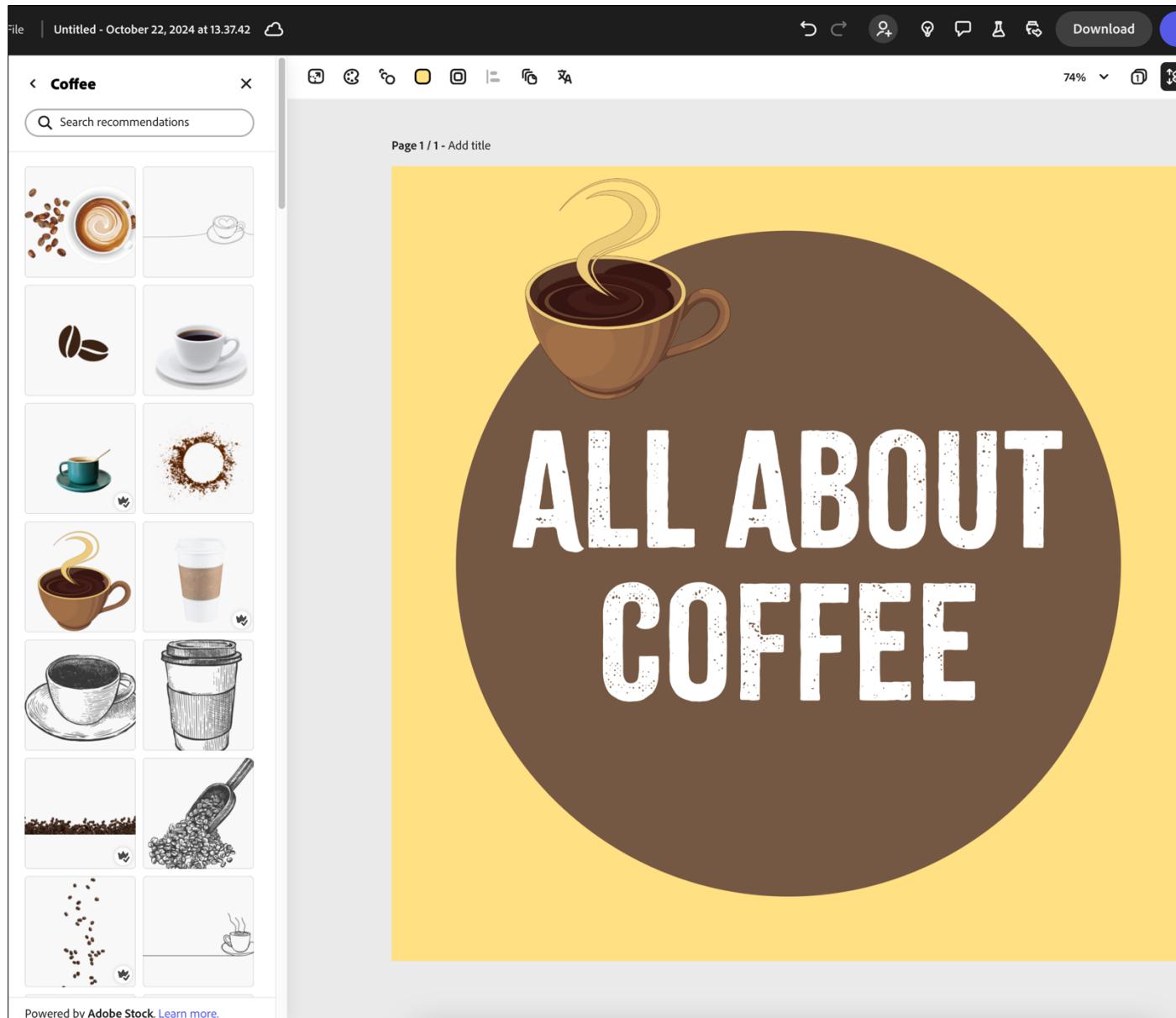
# Recommendation Evaluation

- Step-wise evaluation
  - Project to intent(s)
  - Project intent to recommended intents
  - Asset quality within recommendation
  - Full experience
  - Ethical AI: Biases
- Built UX Add-On to see full experience
- Randomly sampled templates
- User (Adobe-internal) creations



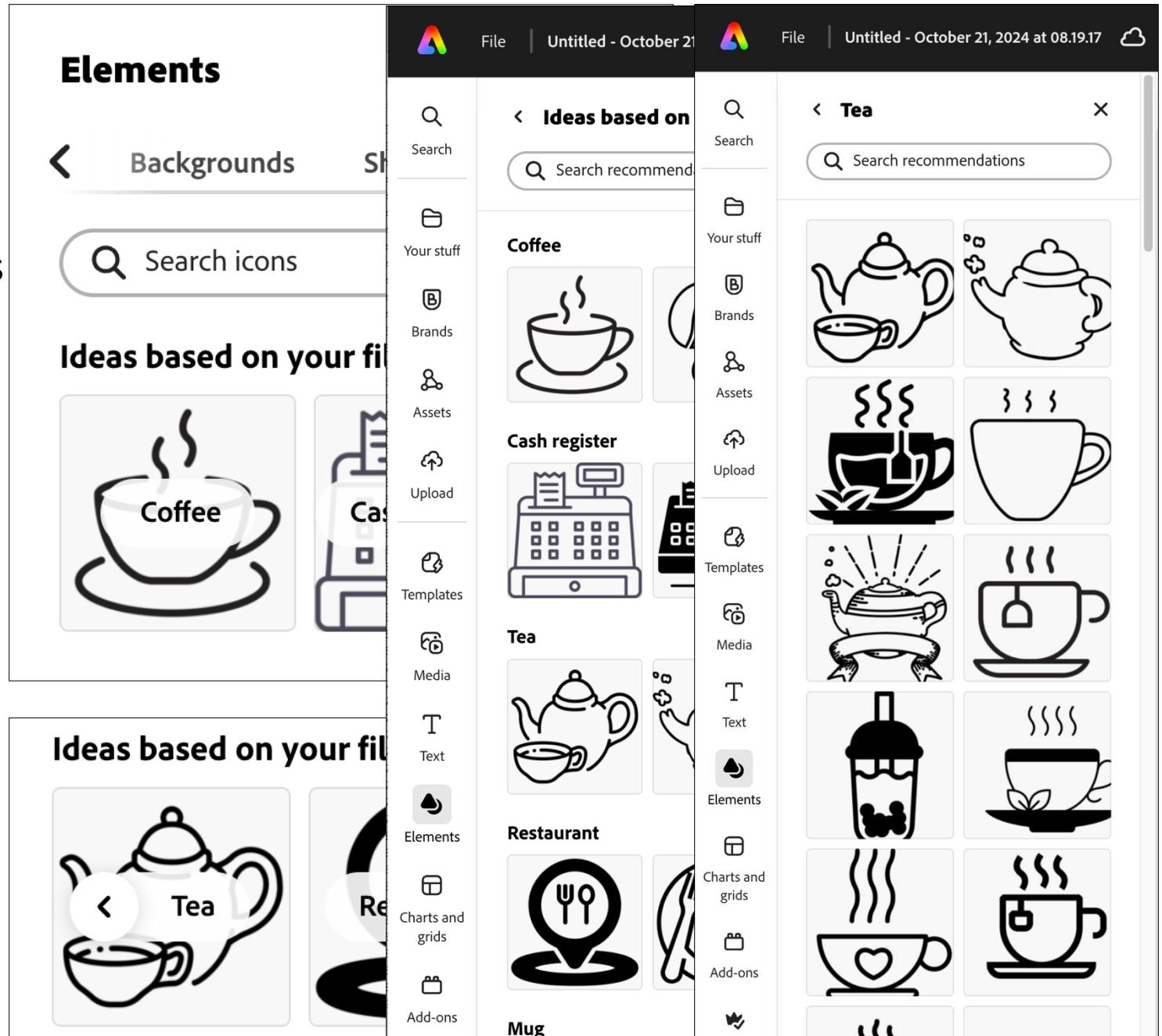
# Evaluation Learnings

- Near duplicates:  
"birthday party" ~ "birthday celebration"
- Bases and associations (e.g. religion)
- Top 3 asset quality is key
  - Issues with search relevance
  - Style and color are sometimes off
- What people want from recs are not always contextually related:
  - visual embellishments
  - shapes (e.g. arrows)
- A canvas has 2-3 intents.
  - Not just “Birthday”: but “Dinosaur”+“Birthday”



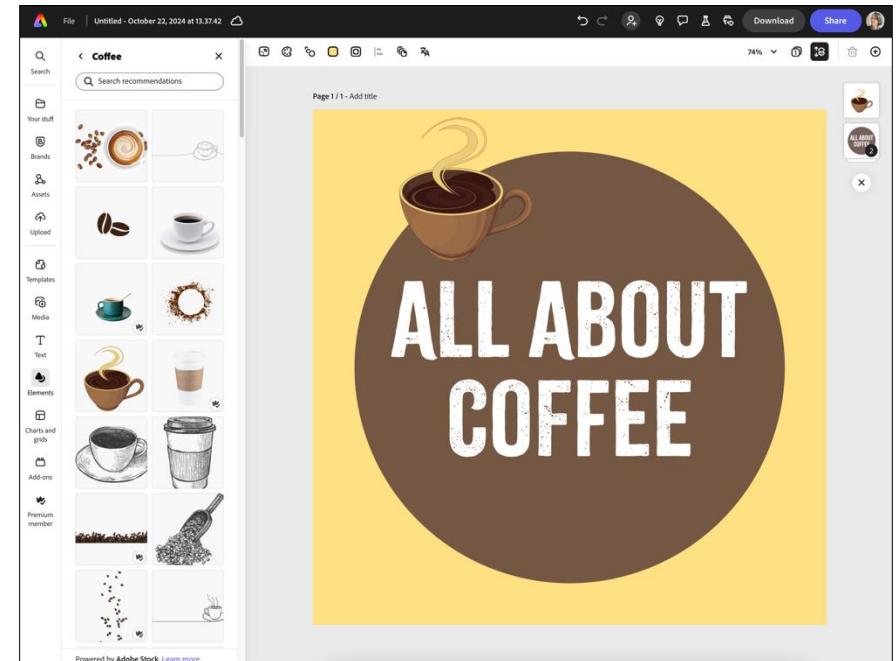
# UX Design

- New top carousel
- Cycles through recommended intents
- Show image as background
- "View all" to see more all intents
- "View all" to see all assets



# Next Steps and Conclusion

- AB testing in progress: Clicks on carousel 1 are up; Export after click on recs is up
- UX iterations
  - Initial carousel optimization
  - Replace "frozen" recommendation carousels
- Algorithm
  - Improved thresholds
  - Style matching for asset results
  - Beyond English
    - Project intent for non-English text
    - Localized display names
    - Style preferences by locale (search ranker)



## References

- Smart Multi-Modal Search: Contextual Sparse and Dense Embedding Integration in Adobe Express  
<https://arxiv.org/abs/2408.14698> (CIKM MMSR 2024)
- Contextual Font Recommendations based on User Intent  
<https://arxiv.org/abs/2306.08188> (SIGIR ECOM 2023)
- Augmenting Knowledge Graph Hierarchies Using Neural Transformers  
<https://arxiv.org/abs/2404.08020> (ECIR 2024)
- Semantic In-Domain Product Identification for Search Queries  
<https://arxiv.org/abs/2404.09091> (SIGIR ECOM 2024)
- Recommending Backgrounds Based on User Intent  
<https://www.freepatentonline.com/y2024/0338553.html> (patent summary)





Adobe