Powering Creative Asset Recommendations with Hybrid Multi-Modal Search

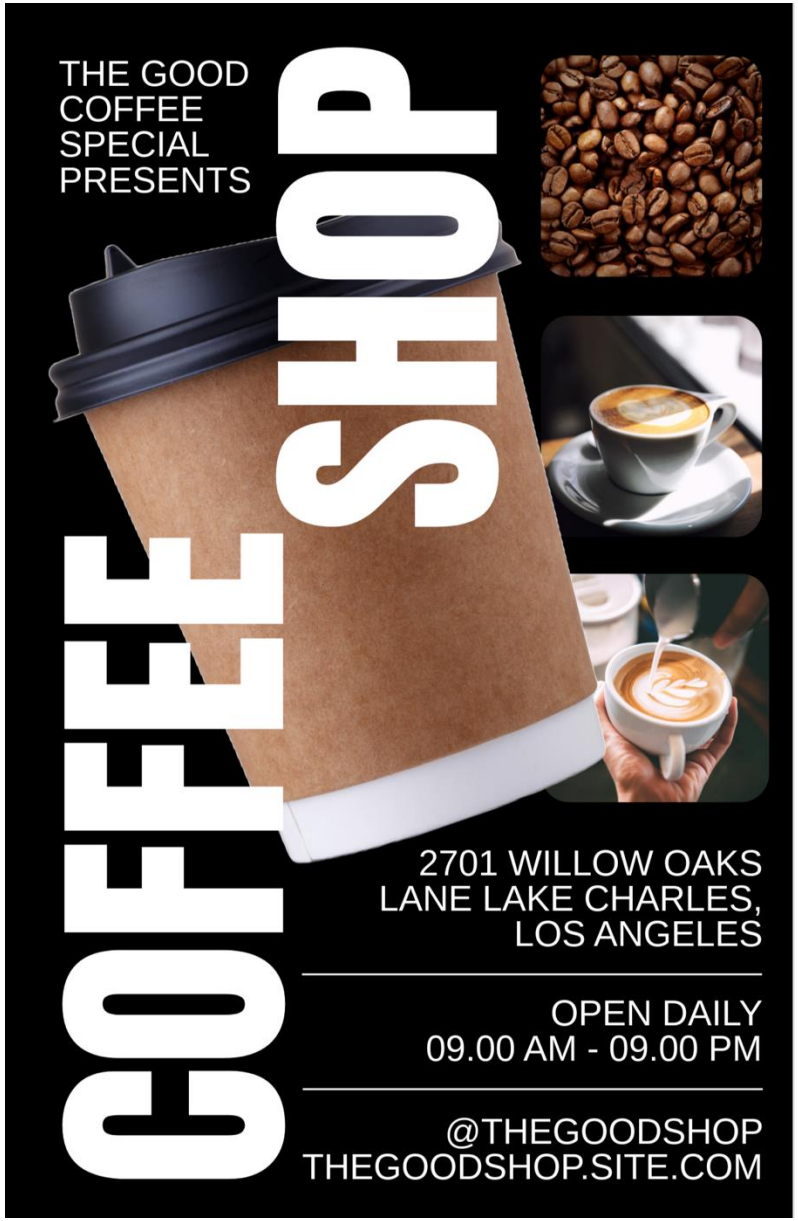
Tracy Holloway King

Adobe

2024-10-25

Intro: Adobe Express Multi-modal Creation

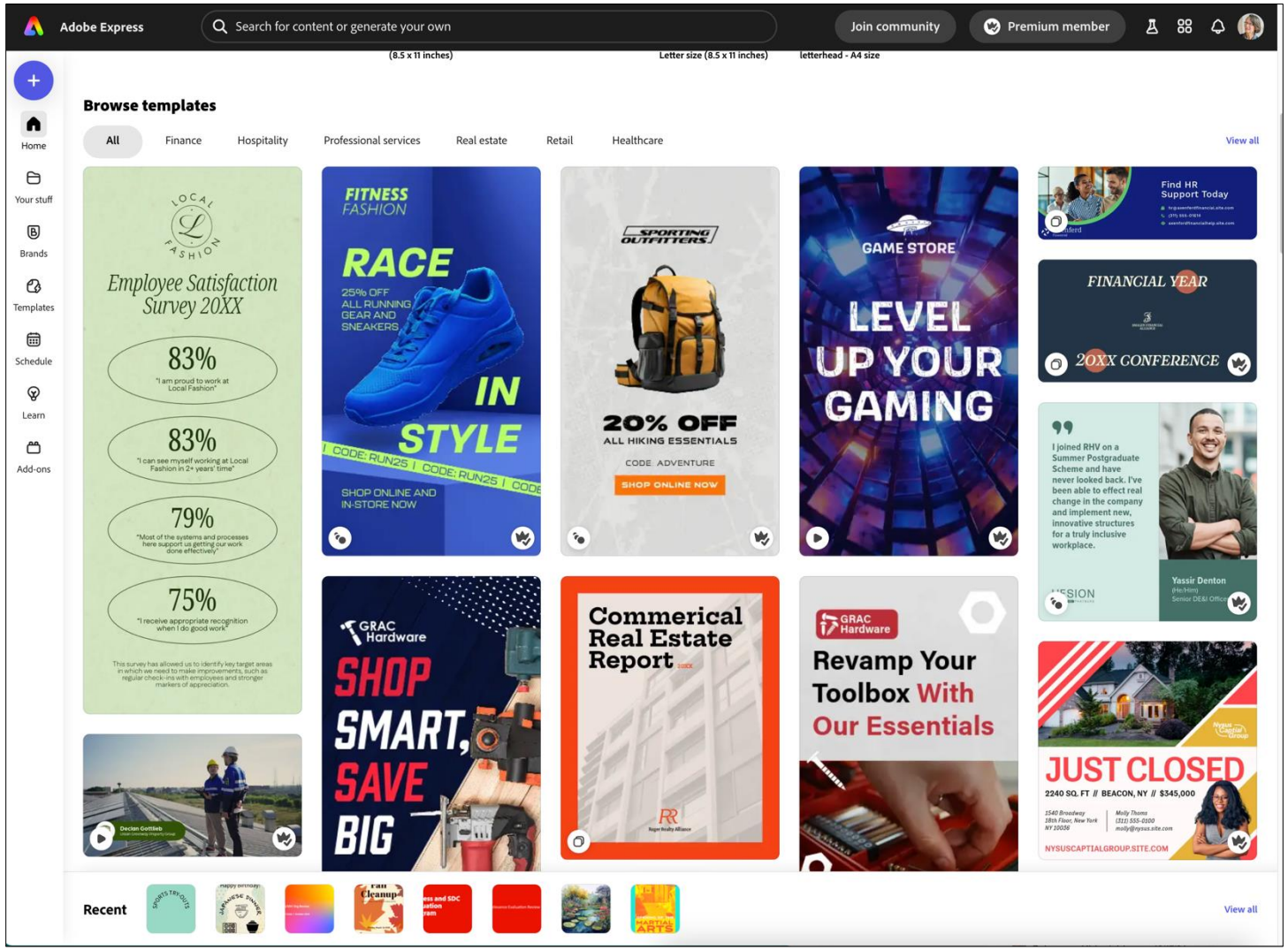
- Creating marketing & social media content is challenging
- Content is multi-modal (multiple images & text)
- Recommendations based on the project:
  - Spark new ideas
  - Discover relevant content
  - Complete projects more quickly
- Contextual recommendation system:
  - Symbolic & embedding intent 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

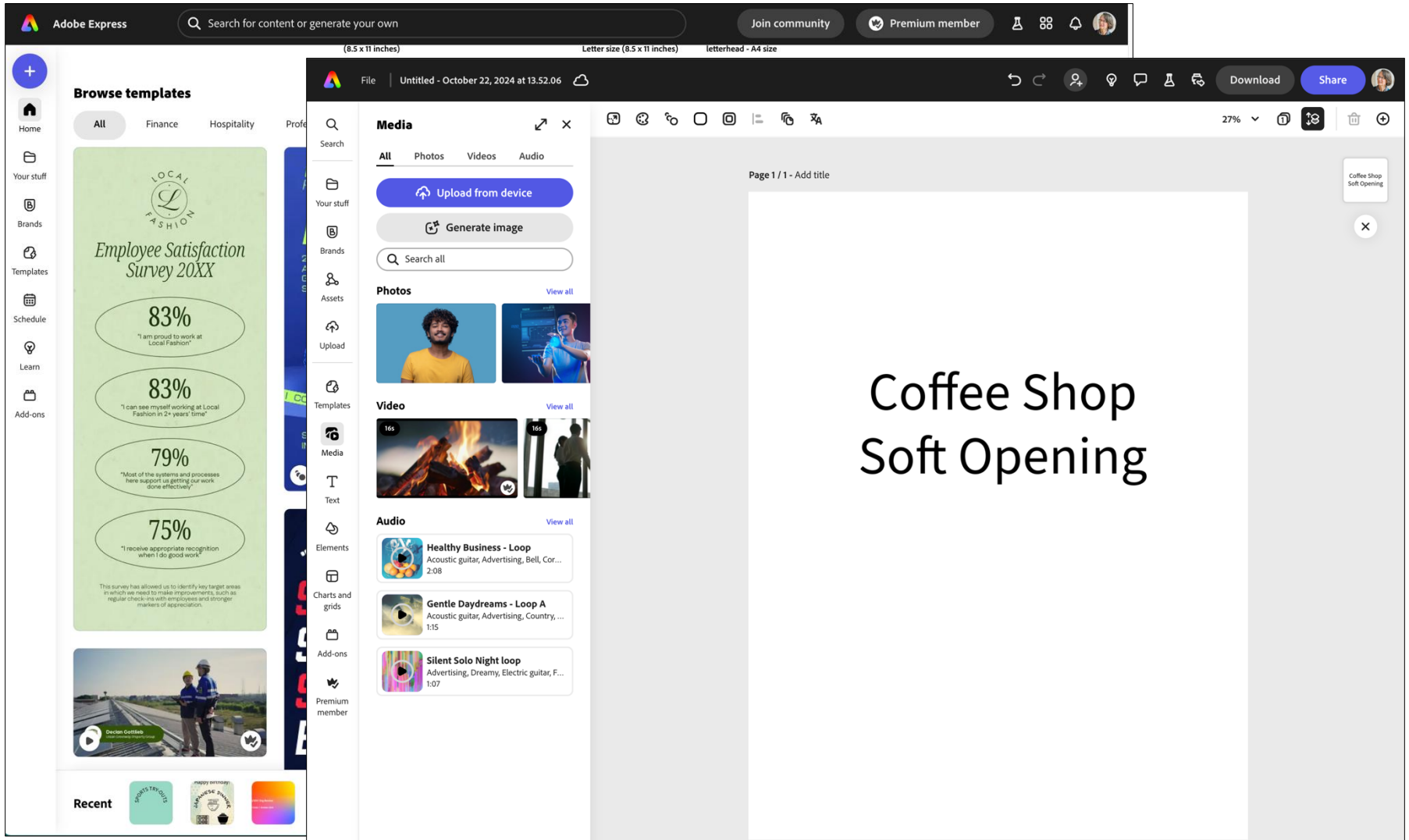
©2020 Adobe. All Rights Reserved. Adobe Confidential.

Adobe Express Project Creation: Start from Template



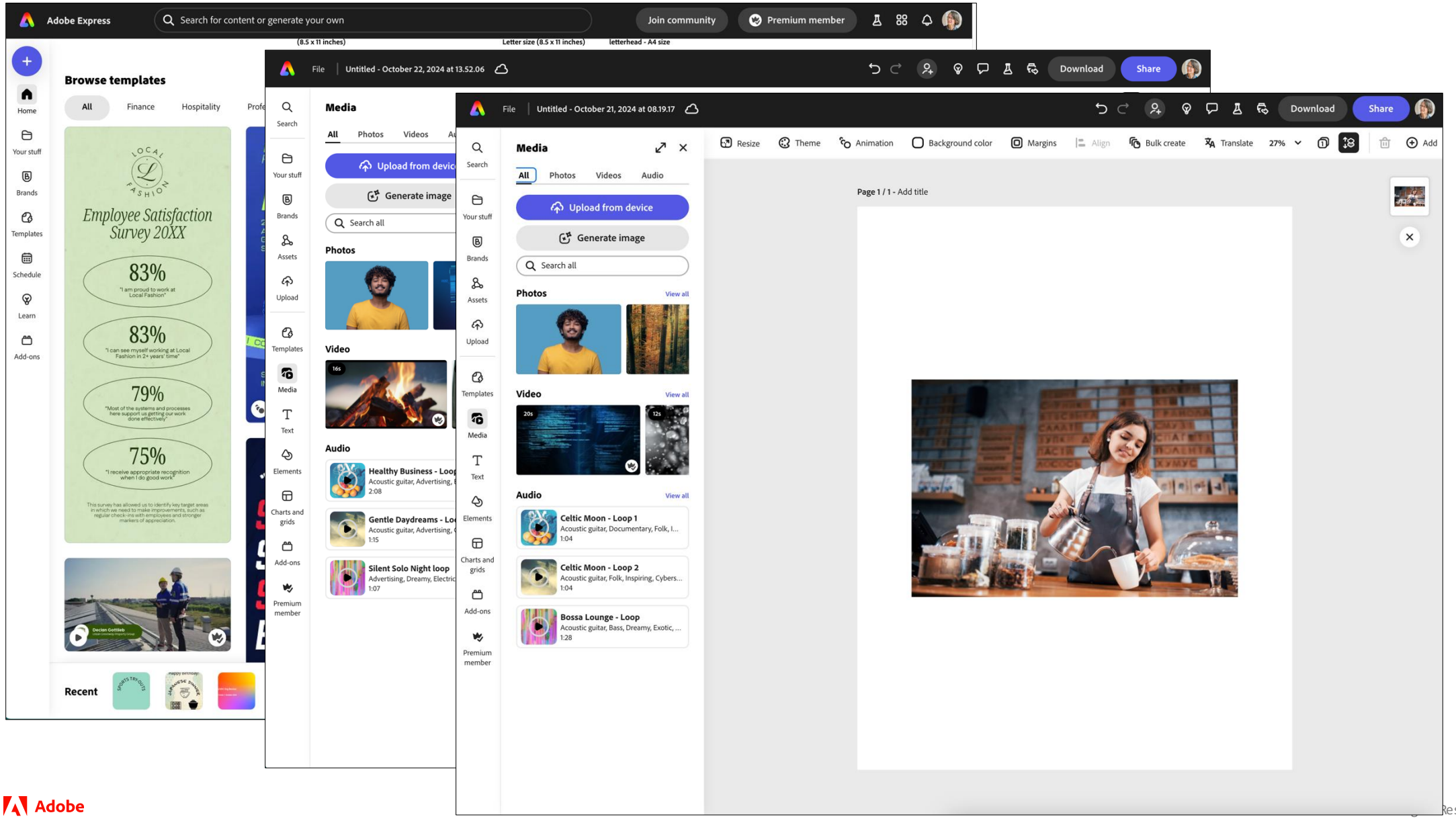
©2020 Adobe. All Rights Reserved. Adobe Confidential.

Adobe Express Project Creation: Start from Text



©2020 Adobe. All Rights Reserved. Adobe Confidential.

Adobe Express Project Creation: Start from Photo



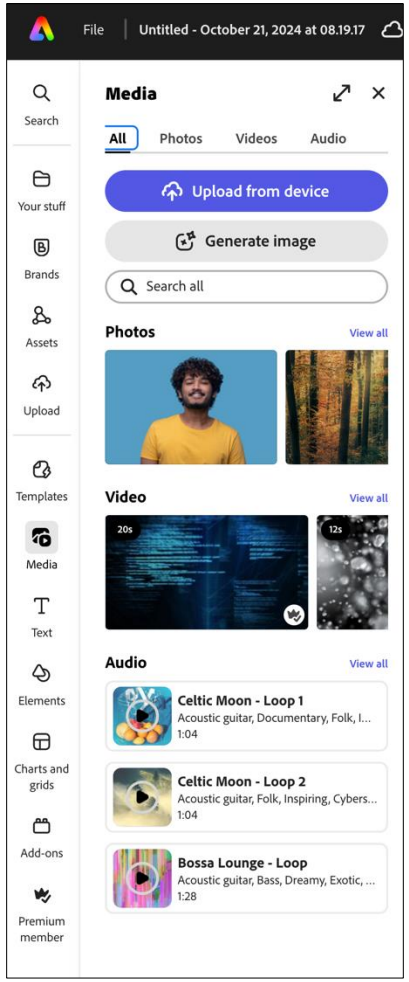
Adobe

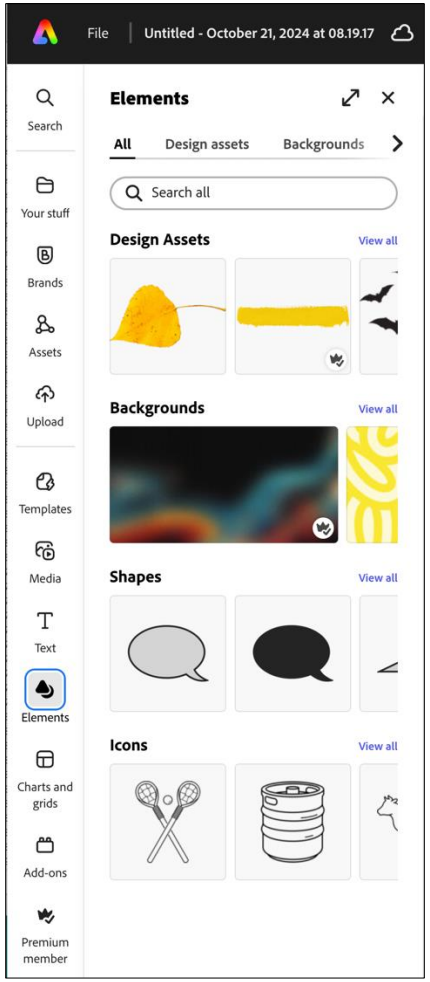
Reserved. Adobe Confidential.

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



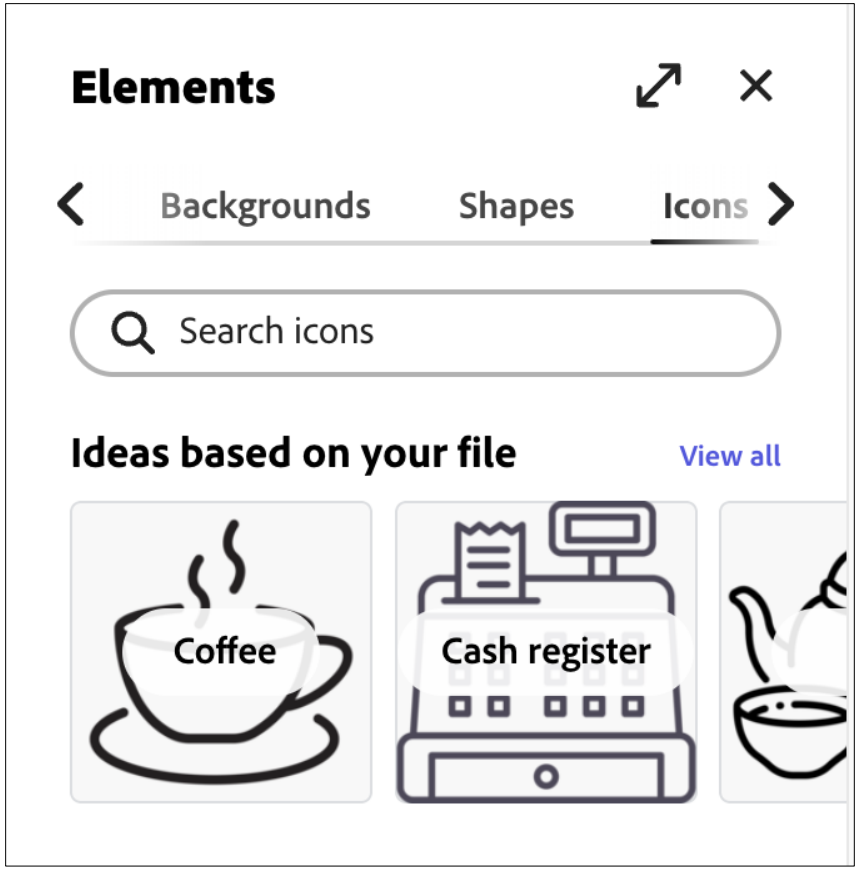


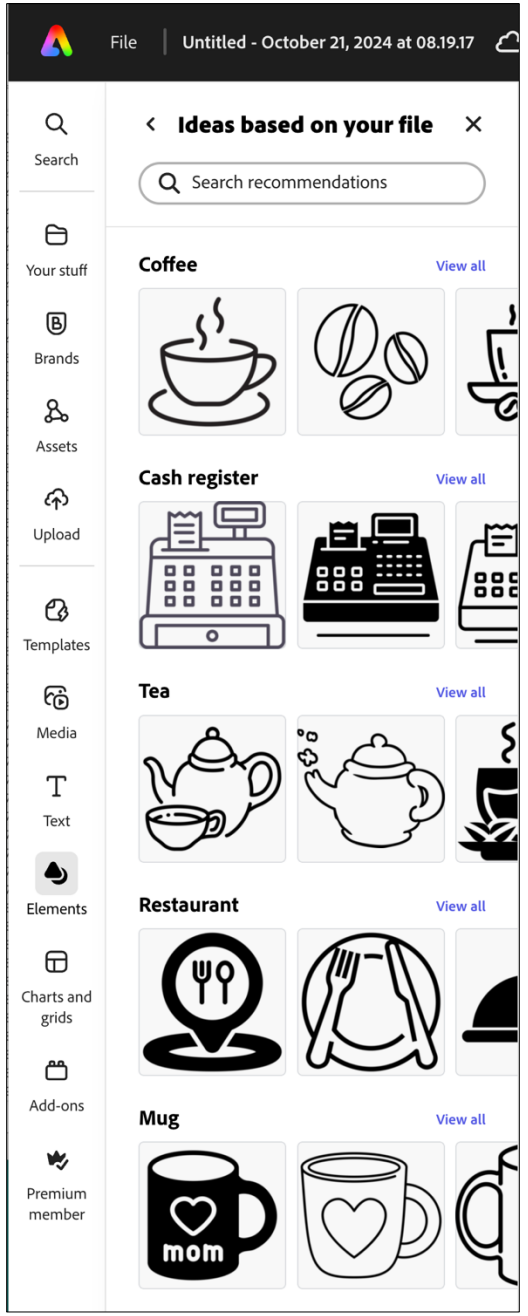
- Goal: Provide asset recommendations based on the user's canvas to improve time to completion.

©2020 Adobe. All Rights Reserved. Adobe Confidential.

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





Adobe

©2020 Adobe. All Rights Reserved. Adobe Confidential.

1. Canvas rendition and inner text are sent to MINT

2. MINT infers the intent of the canvas

MINT

Creative topics
1. Cafe
2. Grand opening
3. Promotion

Associations
1. Barista
2. Mug
3. Latte art
4. Discount tag
5. Open sign
6. Pour Over

Element Queries
- Photos
- Icons
- Design Assets
- Backgrounds

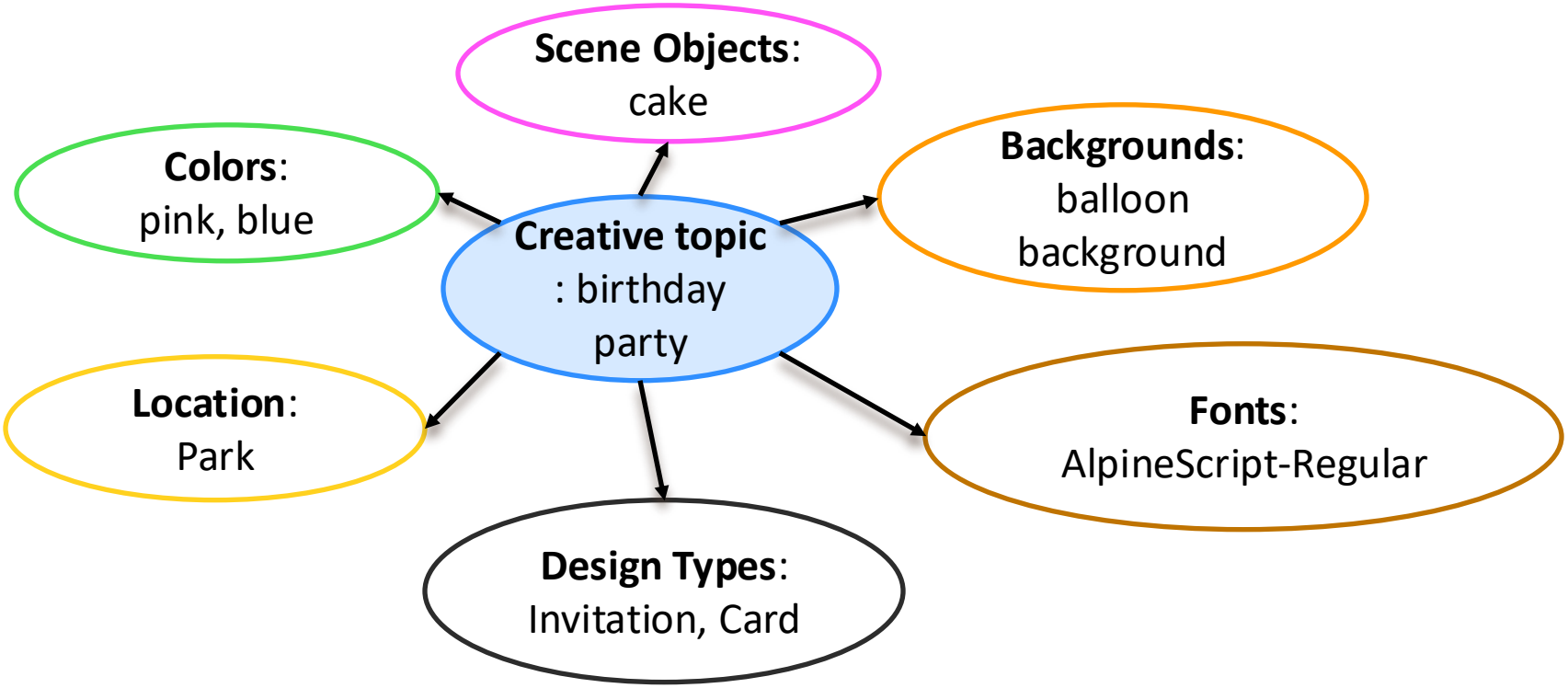
3. MINT looks for the related associations around these topics

4. Asset queries are sent to USS

Adobe
©2020 Adobe. All Rights Reserved. Adobe Confidential.

Creative Knowledge Graph

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



Adobe

©2020 Adobe. All Rights Reserved. Adobe Confidential.

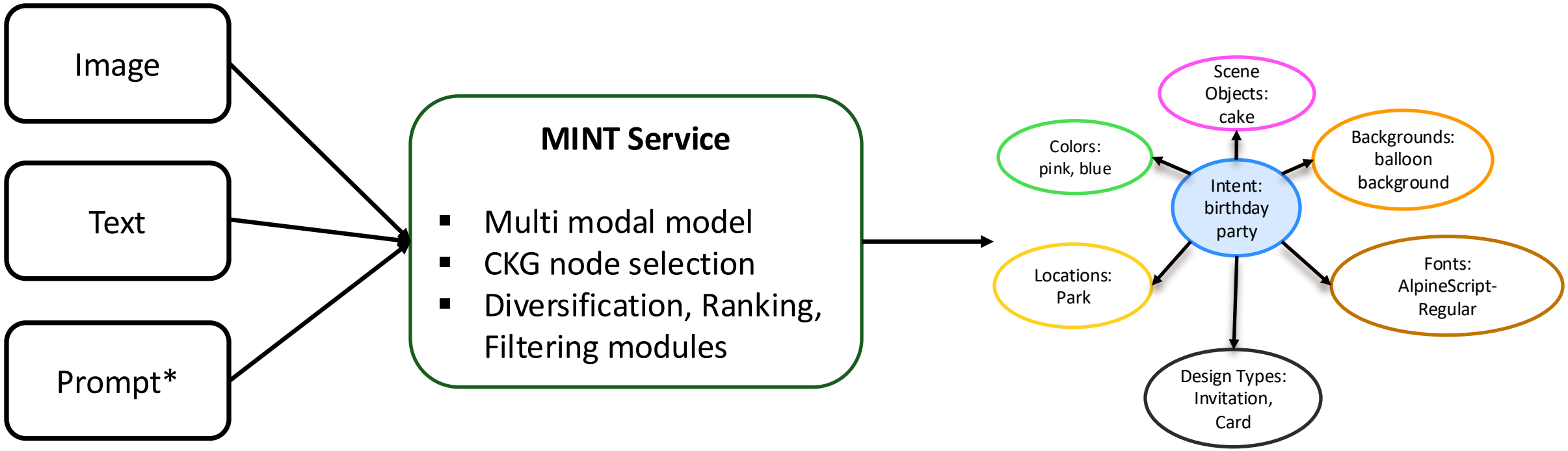
Creative Knowledge Graph

- 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

©2020 Adobe. All Rights Reserved. Adobe Confidential.

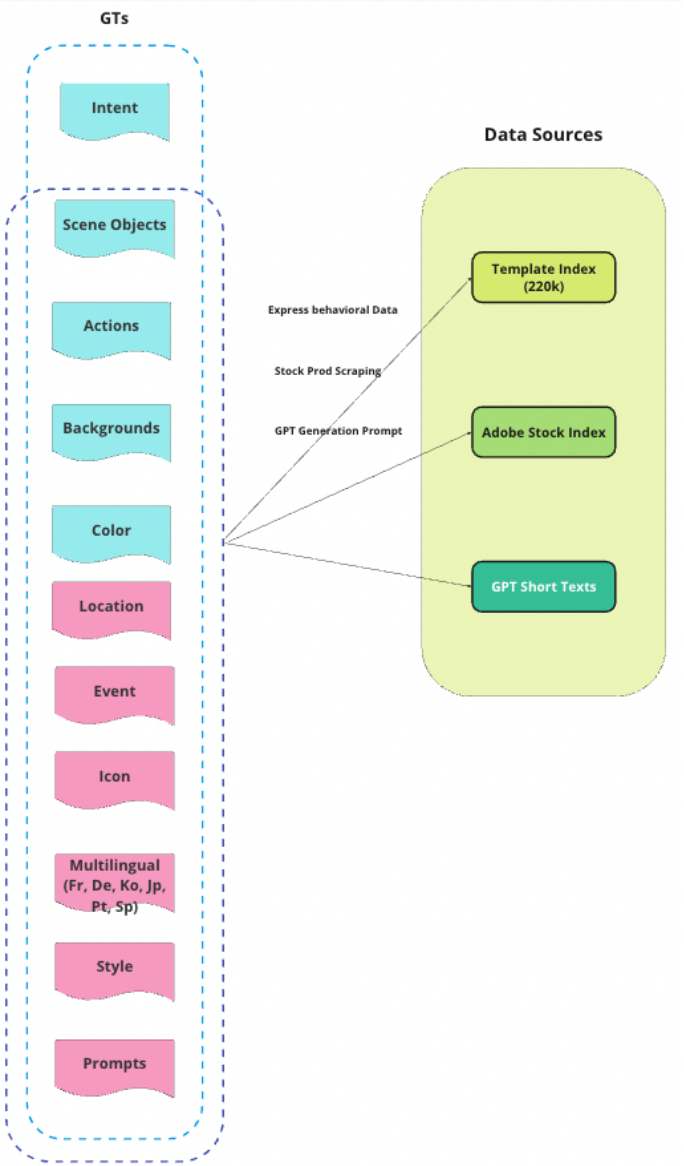
Intent Understanding

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



Adobe
©2020 Adobe. All Rights Reserved. Adobe Confidential.

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 ** only texts*
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) ** translated GPT Texts + Stock Images*
TOTAL	~50M	

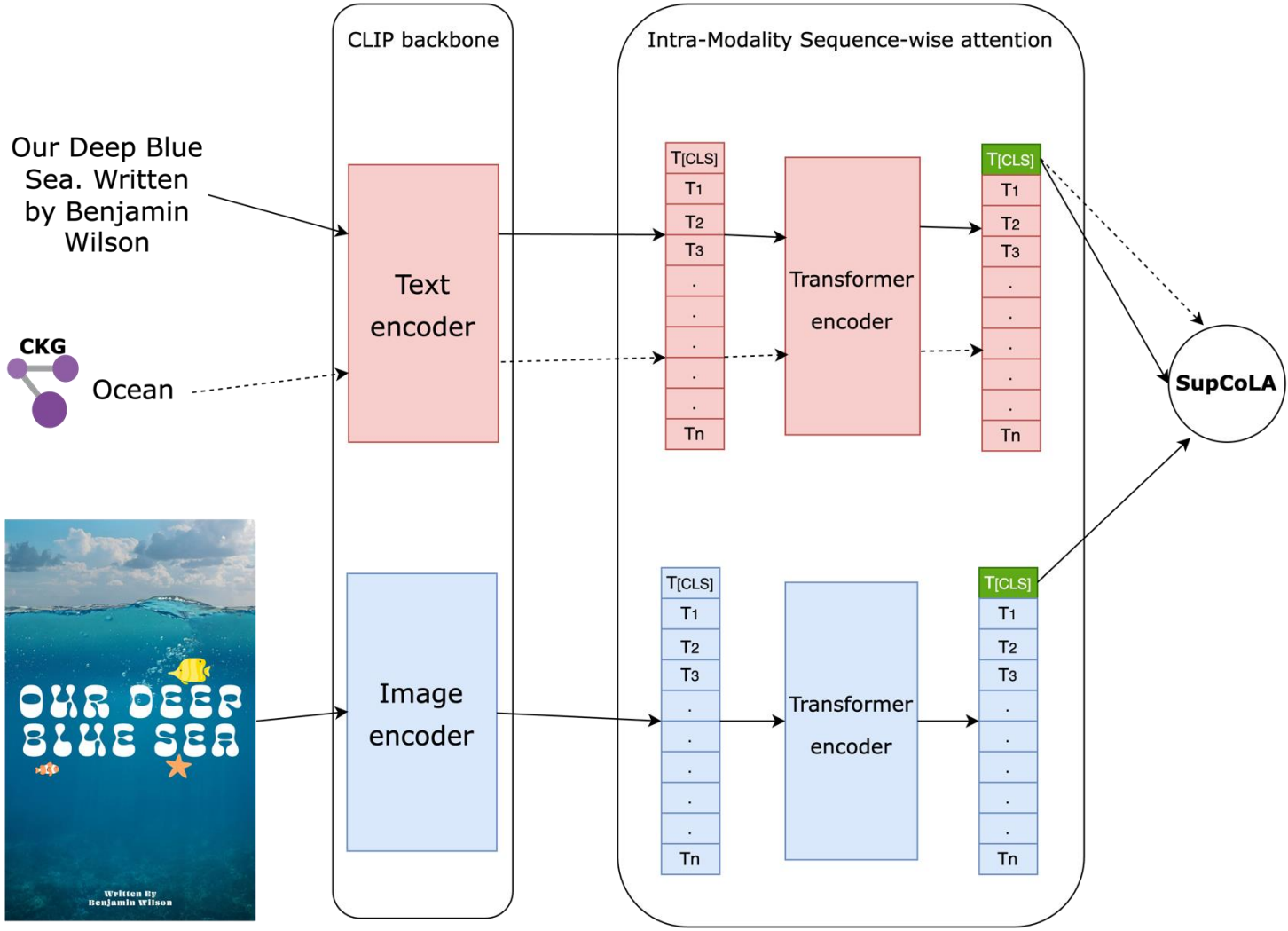
Adobe

©2020 Adobe. All Rights Reserved. Adobe Confidential.

Multi-Modal Model Architecture

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

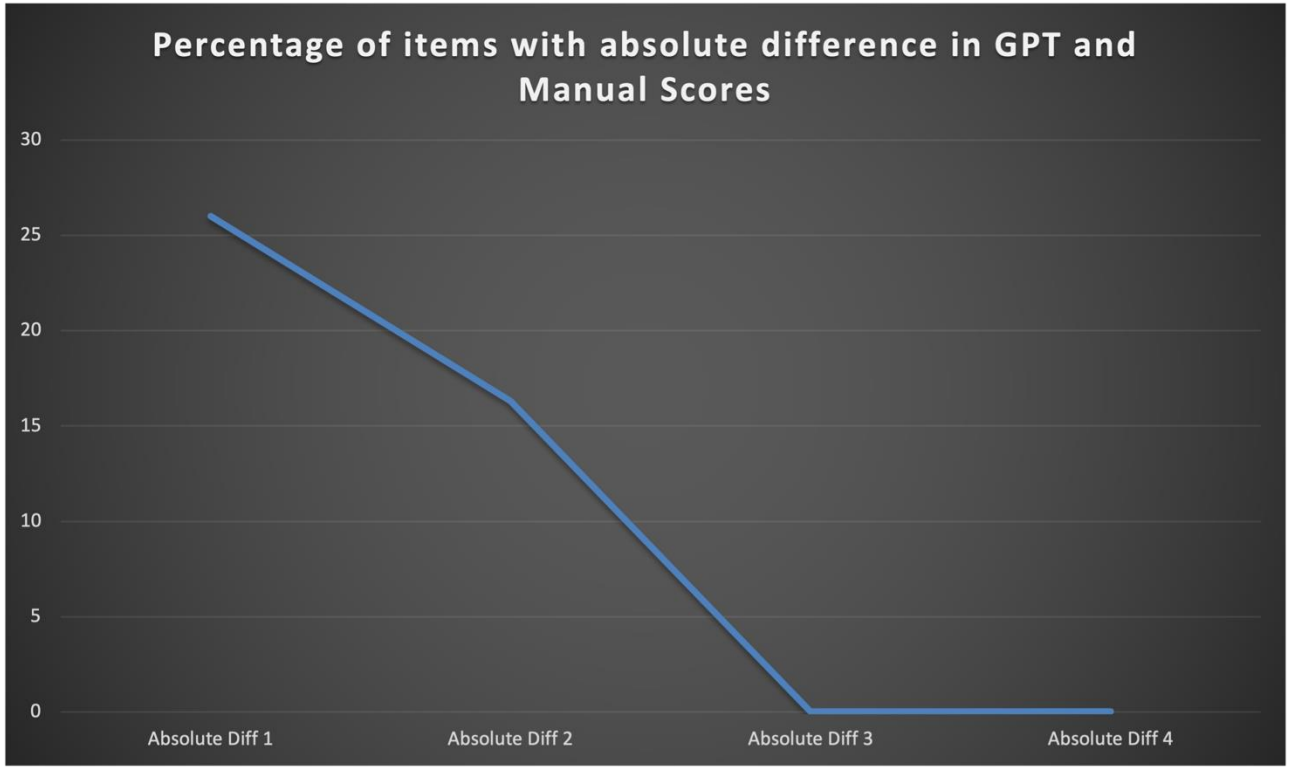
- **Sequence-wise self-attention blocks inspired from CMA-CLIP architecture**
  - Lean multi-head transformer-encoders on top of CLIP backbone.
  - Applies self-attention to individual modality
  - Finetune image or text encoders independently based on use cases.



Adobe
©2020 Adobe. All Rights Reserved. Adobe Confidential.

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



Adobe
©2020 Adobe. All Rights Reserved. Adobe Confidential.

CKG for Contextual Recommendations

1. Canvas rendition and inner text are sent to MINT

2. MINT infers the intent of the canvas

MINT

Creative topics	1. Cafe 2. Grand opening 3. Promotion
Associations	1. Barista 2. Mug 3. Latte art 4. Discount tag 5. Open sign 6. Pour Over
Element Queries	• Photos • Icons • Design Assets • Backgrounds

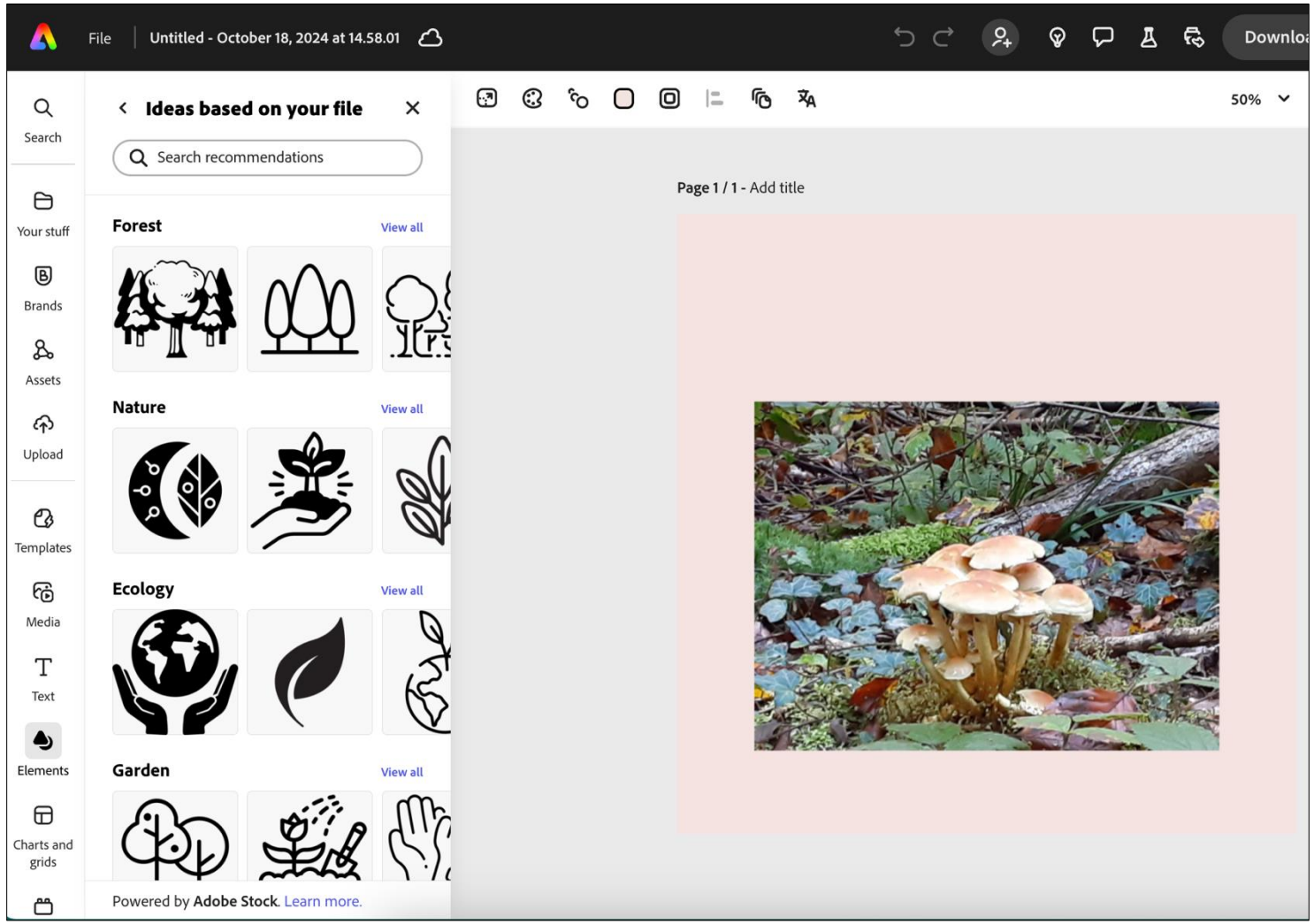
3. MINT looks for the related associations around these topics

4. Asset queries are sent to USS

Adobe
©2020 Adobe. All Rights Reserved. Adobe Confidential.

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



Adobe

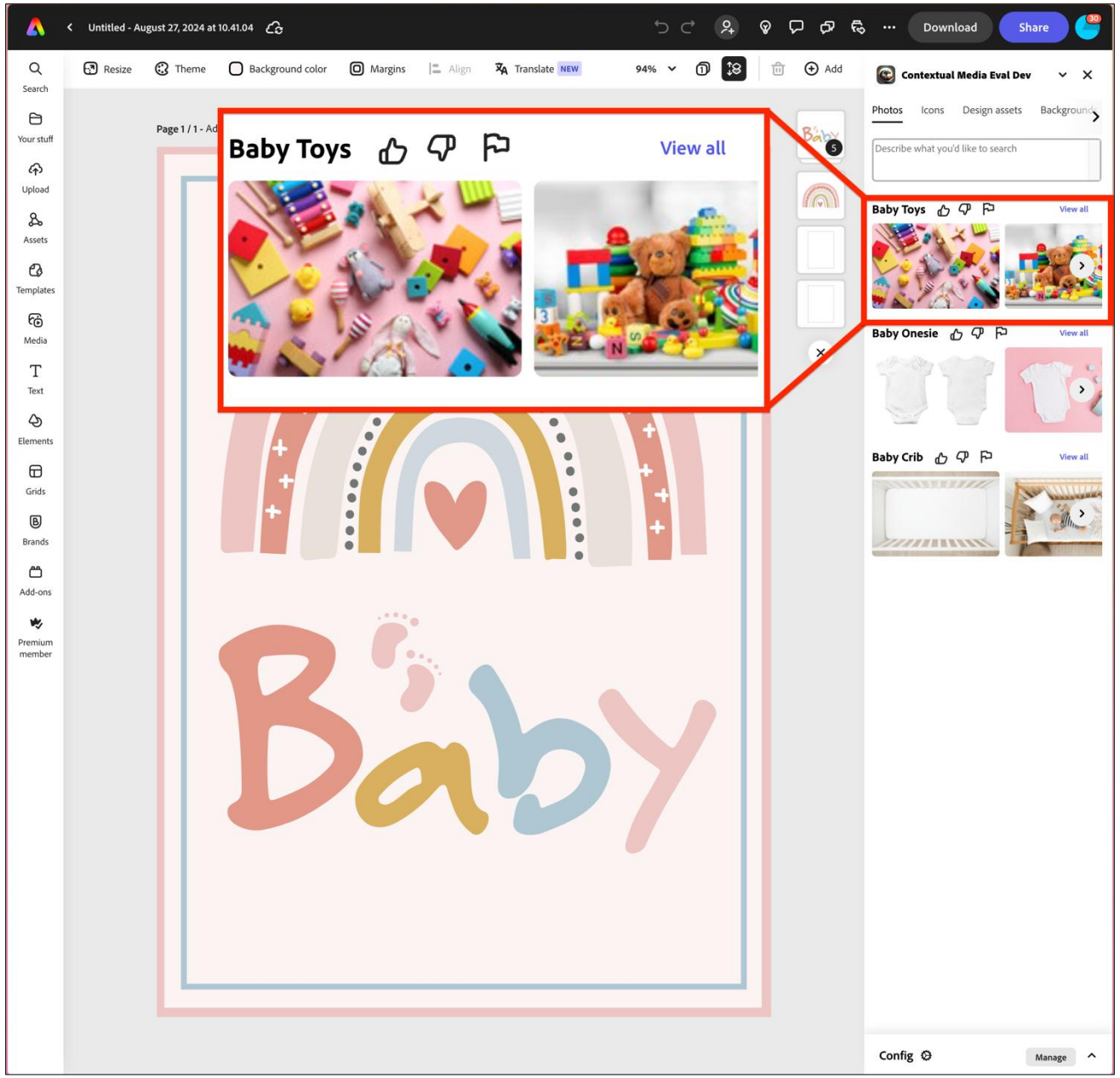
©2020 Adobe. All Rights Reserved. Adobe Confidential.

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



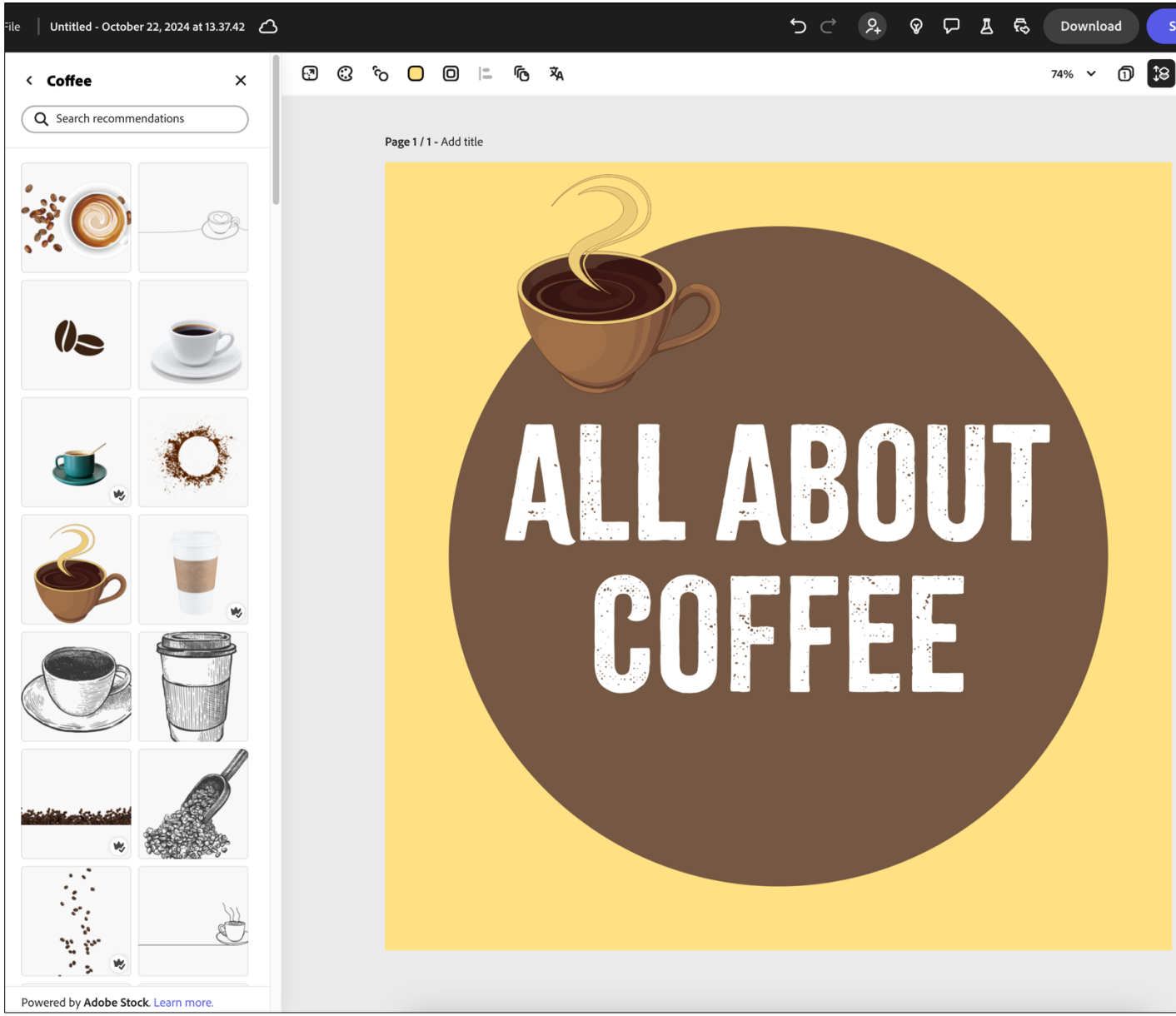
Adobe

©2020 Adobe. All Rights Reserved. Adobe Confidential.

Evaluation Learnings

- Near duplicates:
  "birthday party" ~ "birthday celebration"
- Baises and associations (e.g. religion)
- Top 3 asset quality is key
  - Issues with search relevance
  - Style and color are sometimes off

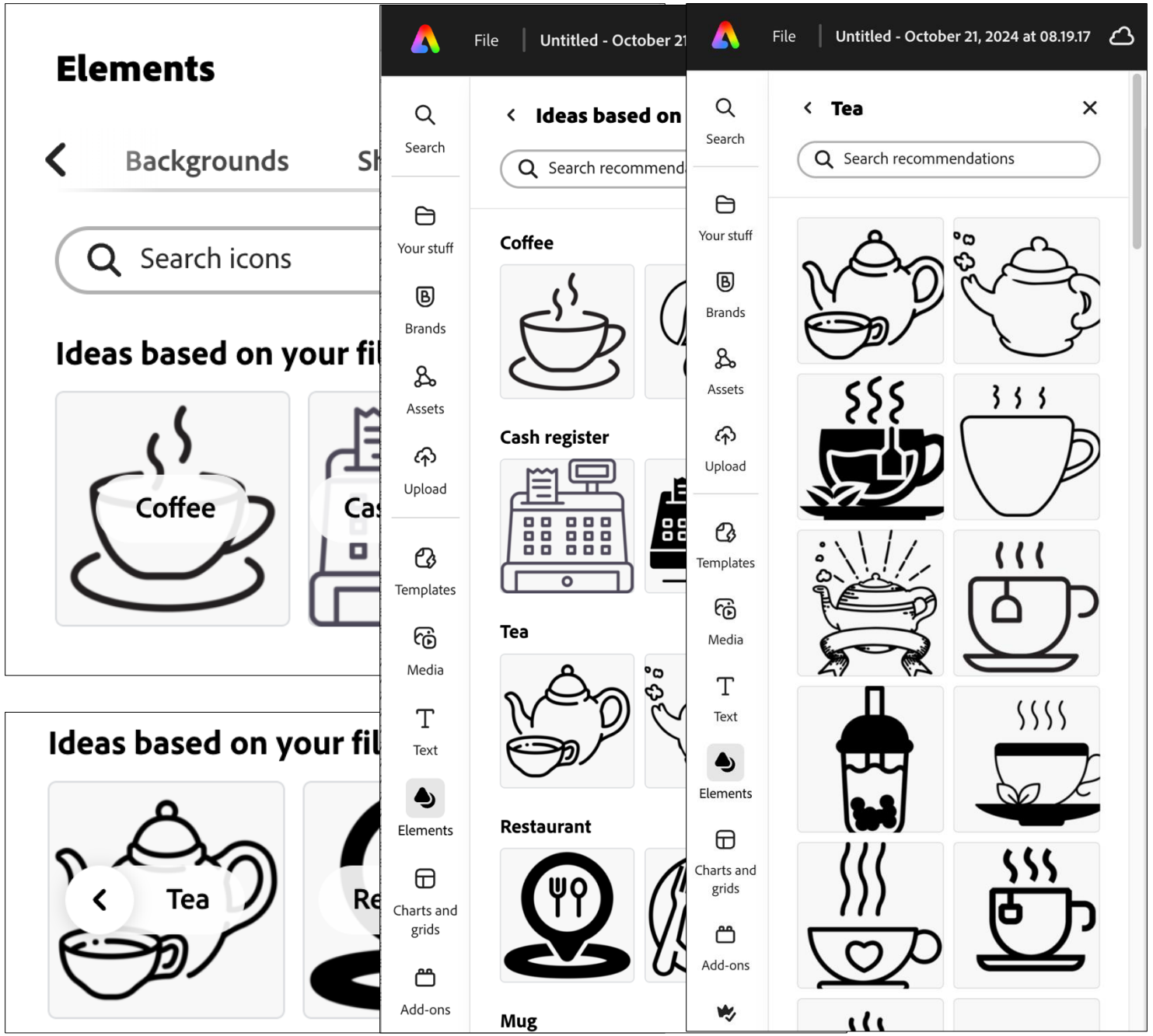
- What people want from recs are not always contextually related:
  - visual embellishments
  - shapes (e.g. arrows)
- A canvas has 2-3 intents.
  - Not just “Birthday”: but “Dinosaur”+“Birthday”



Adobe
©2020 Adobe. All Rights Reserved. Adobe Confidential.

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



Adobe

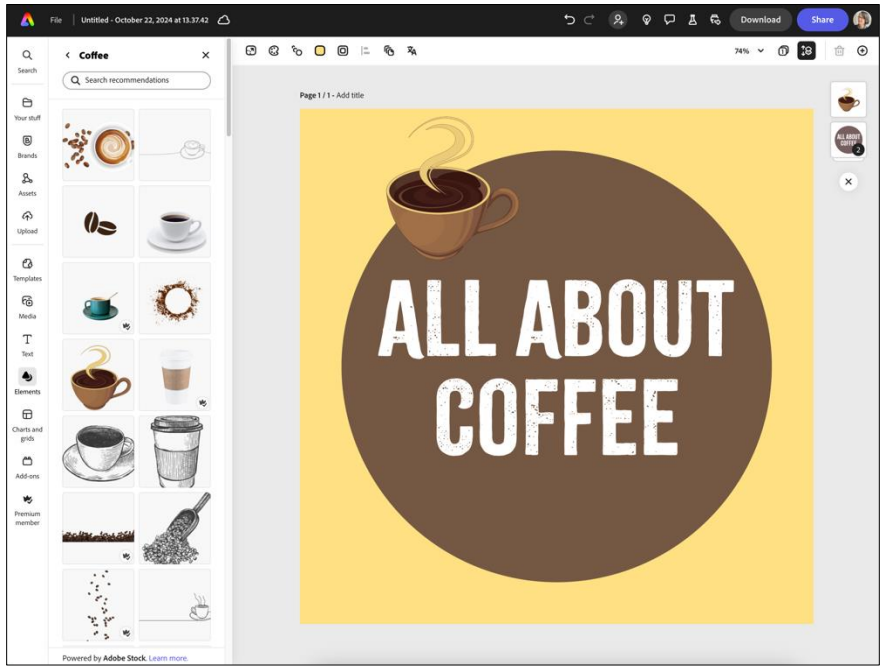
©2020 Adobe. All Rights Reserved. Adobe Confidential.

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)



Adobe
©2020 Adobe. All Rights Reserved. Adobe Confidential.

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.freepatentsonline.com/y2024/0338553.html (patent summary)

Adobe

©2020 Adobe. All Rights Reserved. Adobe Confidential.

Adobe