

Napkinmatic app as a Ubiquitous Pocket AGI: Utility-context-sensitive 3D AR XR HCI for Vision-to-LLM

Yosun Chang
yc@AIMagical.com
AIMagical AI3D.foundation
AReality3D Permute.xyz
San Francisco, CA, USA



Figure 1: Screenshots (left to right): LLM function calling commands, Asking a question about the image, Showing a sample lookup 3D model based on a diagram [Chang 2023f], Contextual-submenu showing options for a venue name in context [Chang 2023h], Toy Ripper Mode [Chang 2023g], product image to 3D, or image to 3D with variable generation times, Before and After images (augmented back) using command image2furni cthululu living room via StableDesign [Lavreniuk 2024]

ABSTRACT

Napkinmatic Ubiquitous is a seamless spatial computing app platform connecting AI to the real world – and back again. Simply take a picture of anything (or import an image), to load an AI Swiss army knife menu that lets you “chat with/narrate the image”, load relevant 3D models, chain models, or AI3D create from image to 3D [Chang 2023e], ControlNet [Zhang et al. 2023] transform the image into a masterpiece painting [Chang 2023b], add interior design to an empty room, and more - and augment it back to the real world.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI); Mixed / augmented reality.**

KEYWORDS

augmented reality, AI, spatial computing, ubiquitous computing

ACM Reference Format:

Yosun Chang. 2023. Napkinmatic app as a Ubiquitous Pocket AGI: Utility-context-sensitive 3D AR XR HCI for Vision-to-LLM. In *Special Interest*

Permission to make digital or hard copies of part or all 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).

SIGGRAPH '23 Appy Hour, August 06-10, 2023, Los Angeles, CA, USA

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0156-6/23/08.

<https://doi.org/10.1145/3588427.3595357>

Group on Computer Graphics and Interactive Techniques Conference Appy Hour (SIGGRAPH '23 Appy Hour), August 06-10, 2023. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3588427.3595357>

1 INTRODUCTION

Napkinmatic [Chang 2023c] [Chang 2023e] is a portmanteau of the word Napkin and automatic. Similar in name and in the spirit of DrawmaticAR [Chang 2020a] [Chang 2020b], it's about extending the potential of your napkin sketch – “automagically.”

We extend our Napkinmatic app to include additional AI models, such as vision-based LLM open prompt image analysis, chain-able image to image, and robust image to 3D. Now we have the entire world and anything in it, as our napkin canvas.

Simply take a picture of anything (or import an image), to load a “utility-context-sensitive” AI Swiss army knife menu that lets you...

2 TEXT INTERACTION

2.1 Chat with the image

You can ask the image anything by chatting with it. If it's a diagram, ask for visual details like how many LEDs does it have?

2.2 Narrate the image

LLMs are capable of impersonating anyone well known now - and in history. For example, we have wandered around an academic conference to have Issac Newton AI pontificate various science-field

posters, or to have Emad himself comment on posters at NeurIPS [Chang 2023a].

3 CONTEXTUAL SUB-MENUS

When specific text-in-context such as a venue name is recognized and selected, a contextual sub-menu can help display the most likely options – load the map, share, or buy items from the venue [Chang 2023h].

4 AI3D METHODS

5 LOAD RELEVANT 3D MODELS IN A DIAGRAM (OR CREATE A NEW ONE)

For a diagram that is standardized, such as Boston Robotics Spot, the model can understand it and search a known database of 3D models [Chang 2023f]. When it isn't recognized, we can use image to 3D and controlnet to AI3D create a new AI-generated 3D model based on the available data [Chang 2023e].

6 INVOKE IMAGE TO 3D

Napkinmatic also does object detection to send sub-images to 3D for robust image to 3D operations - from ripping 3D printable toys from a toy product image [Chang 2023g] to segmenting a photo into identified objects and using z-depth estimations to create 3D models [Chang 2024a] to letting a user specify the image object they want to transform [Chang 2023e].

7 CHAIN MODELS TO PROCESS THE INPUT IMAGE

Extending the concept of turning a sketch or an outline to a vivid rendering, we now have multiple models. We simply chain the models together (or undo) to create new variations [Chang 2023e].

8 INPAINT/OUTPAINT

Modify the image or extend the image.

9 ADD INTERIOR DESIGN TO AN EMPTY ROOM

Models such as Stable Design [Lavreniuk 2024] let you add real world objects such as furniture to an image of an empty room - in line with interior design style constraints.

10 VIRTUAL TRY-ON

Take a photo of anyone to generate try-on images of any clothing.

11 PREVIOUS FEATURES

ControlNet [Zhang et al. 2023] transform the image into a masterpiece painting [Chang 2023b], and transforming sketches to different image styles via Control-Net [Zhang et al. 2023] and doing image to 3D [Chang 2023e] with rigged dlib-defined characters.

11.1 API Response Data Format

When Napkinmatic was published last year, we established a general API format as defined in the Napkinmatic Github [Chang 2023d].

Since then, the industry has a lot of fragmentation, and so we have simply adapted the differences in each different model or third party service, to create our own universal API, such as the AI3D API [Chang 2024b].

11.2 When anything can become an interface, make it so.

Napkinmatic places computing on any surface in reality, and thus, we can turn any surface into a computing platform.

11.3 And when it doesn't make sense to be in reality...

We can always step back to simply allowing the user to load an image in non-AR mode.

12 CONCLUSION: TOWARDS TRANSPARENT USER INTERFACES

The best UI is transparent – a user experience that is so naturally intuitive, it basically disappears. Barring that, it needs to not hamper the user with additional learning curves or unnecessary tedium time in needing to use interfaces that do not scale. We believe that with the transformative power of AI and the many more parameters of control yet to come, we can achieve the fully transparent user interface.

A quote from [Chang 2023b] last year - although our form factor here is a phone or tablet, we are definitely on our way there!

REFERENCES

- Yosun Chang. 2020a. *Special Interest Group on Computer Graphics and Interactive Techniques Conference Real-Time Live! (SIGGRAPH '20 Real-Time Live!)* (Aug. 2020). <https://doi.org/10.1145/3407662.3407755>
- Yosun Chang. 2020b. DrawmaticAR: automagical AR content from written words!. In *ACM SIGGRAPH 2020 Real-Time Live!* (Virtual Event, USA) (SIGGRAPH '20). Association for Computing Machinery, New York, NY, USA, Article 3, 1 pages. <https://doi.org/10.1145/3407662.3407755>
- Yosun Chang. 2023a. #EmadAI via SMBDY.ai. <https://x.com/Yosun/status/1735798618328584438>.
- Yosun Chang. 2023b. Napkinmatic. In *ACM SIGGRAPH 2023 Appy Hour* (Los Angeles, CA, USA) (SIGGRAPH '23). Association for Computing Machinery, New York, NY, USA, Article 4, 2 pages. <https://doi.org/10.1145/3588427.3595357>
- Yosun Chang. 2023c. Napkinmatic. In *ACM SIGGRAPH 2023 Appy Hour* (Los Angeles, CA, USA) (SIGGRAPH '23). Association for Computing Machinery, New York, NY, USA, Article 4, 2 pages. <https://doi.org/10.1145/3588427.3595357>
- Yosun Chang. 2023d. *Napkinmatic - Open Framework*. Retrieved May 23, 2023 from <https://github.com/yosun/napkinmatic>
- Yosun Chang. 2023e. Napkinmatic 3D. In *ACM SIGGRAPH 2023 Real-Time Live!* (Los Angeles, CA, USA) (SIGGRAPH '23). Association for Computing Machinery, New York, NY, USA, Article 6, 2 pages. <https://doi.org/10.1145/3588430.3597253>
- Yosun Chang. 2023f. Napkinmatic TED AI expo demo. <https://x.com/Yosun/status/1714437768502534423>.
- Yosun Chang. 2023g. Napkinmatic3D Toy Ripper Mode. <https://x.com/Yosun/status/1721892003041149434>.
- Yosun Chang. 2023h. PocketAGI Hackathon App. <https://devpost.com/software/pocketagi>.
- Yosun Chang. 2024a. AI3D Camera Obscura. <https://x.com/Yosun/status/1774521401355166148/>.
- Yosun Chang. 2024b. VolumeMatic and the AI3D.foundation: Text to interactive volume app spatial computing creator app. *SIGGRAPH '24: ACM SIGGRAPH 2024 Appy Hour* (2024). <https://dl.acm.org/doi/10.1145/3664294.3664361>
- Ludwiczuk Bartosz Lavreniuk, Mykola. 2024. StableDesign. <https://github.com/Lavreniuk/generative-interior-design>.
- Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models.