SAGIVTECH

IMVC 2024

Kiki or Bouba? Sound Symbolism in Vision-and-Language Models

Morris Alper and Hadar Averbuch-Elor

NeurIPS 2023 ✨SPOTLIGHT✨

VAILab

TEL AVIV UNIVERSITY אוניברסיטת תל אביב

SAGIVTECH
IMVC 2024

Next you will be shown two images that were generated by Stable Diffusion using the prompt:

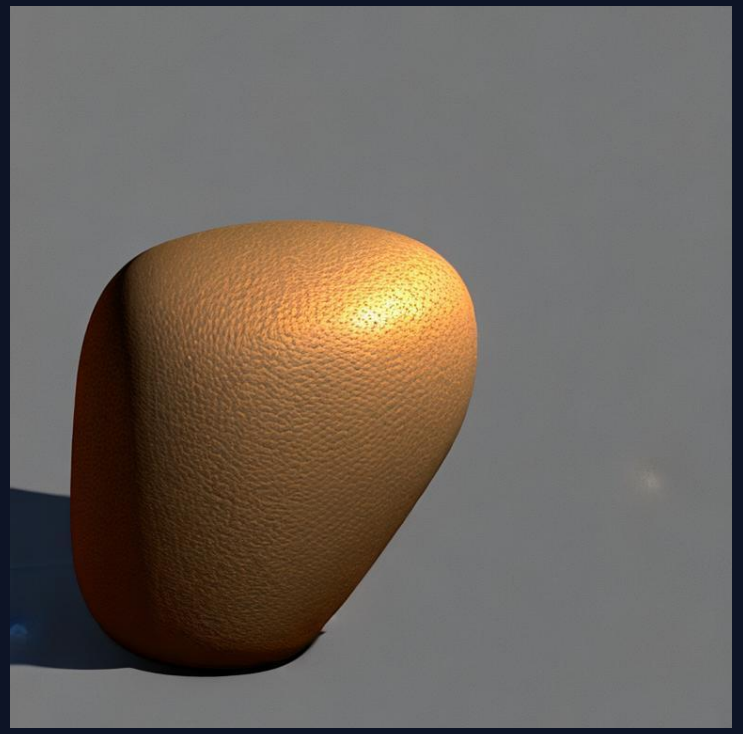
"A 3D rendering of a _____-shaped object"

For one the missing word is *bouba* and for the other *kiki*.

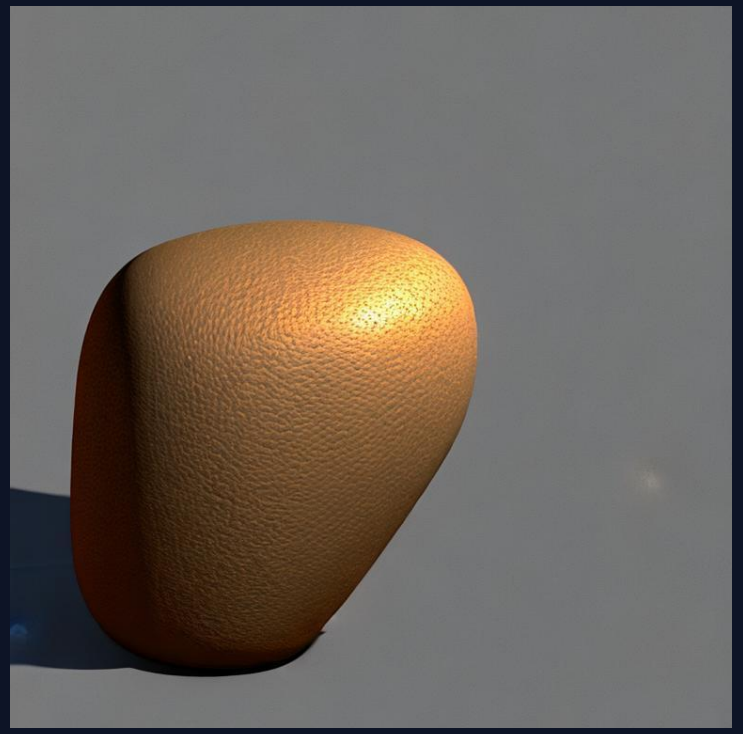
Which is which?

SAGIVTECH

IMVC 2024



SAGIVTECH
IMVC 2024



bouba

kiki

The Kiki-Bouba Effect

A well-known study in psycholinguistics

Holds across cultures, languages, and even among infants.

[Blasi et al. 2016, Ćwiek et al. 2022, Maurer et al. 2006, Ozturk et al. 2013]

SAGIVTECH
IMVC 2024

Sound Symbolism

Connections between sound and meaning in language

kapow, glub-glub, boom

crunch/crush/crash/crinkle/crack

mama　　אמא (ima)　　妈妈 (māma)

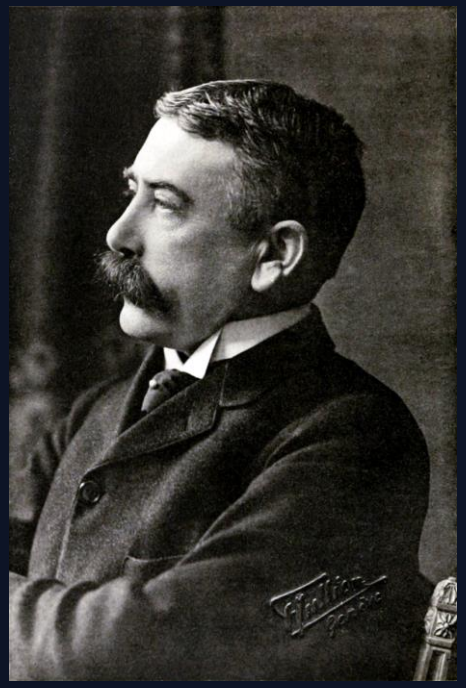
SAGIVTECH
IMVC 2024

Sound Symbolism

Connections between sound and meaning in language

“Le signe est arbitraire” !
(the sign is arbitrary)

Ferdinand de Saussure
(1857–1913)



SAGIVTECH
IMVC 2024

Sound Symbolism

Connections between sound and meaning in language

Have VLMs learned sound symbolism?

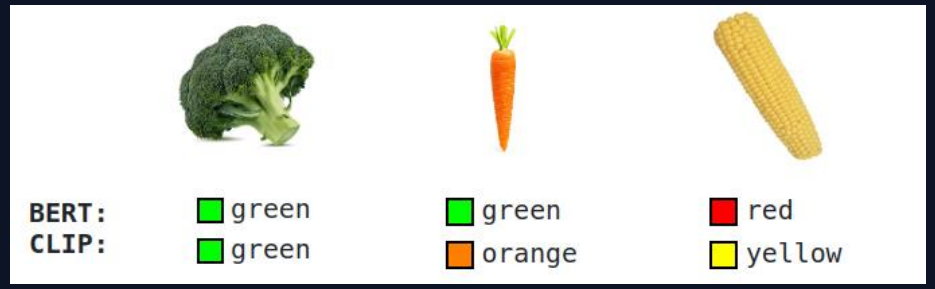
Why is this important?

Interpreting the **black box of foundation VLMs**

Why is this important?

Interpreting the **black box of foundation VLMs**

Tewel et al. (CVPR 2022)

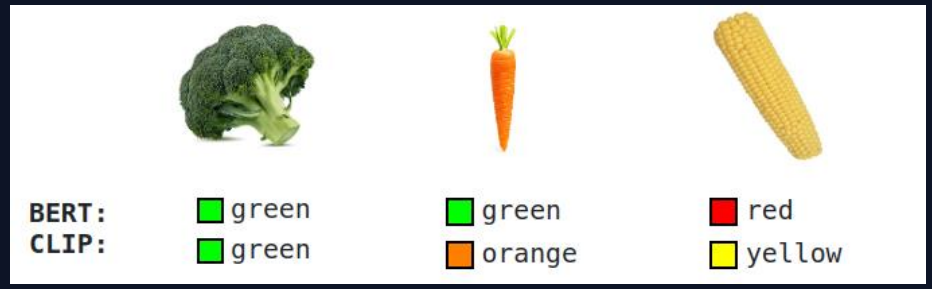


Alper et al. (CVPR 2023)

Why is this important?

Interpreting the **black box of foundation VLMs**

Tewel et al. (CVPR 2022)



Alper et al. (CVPR 2023)

New tool for studying language

Insight into how sound symbolism could be learned

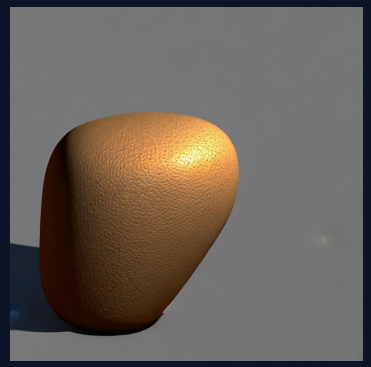
Probing VLMs for Sound Symbolism

SAGIVTECH
IMVC 2024

Probing VLMs for Sound Symbolism

Measuring geometric attributes of pseudowords

rounder

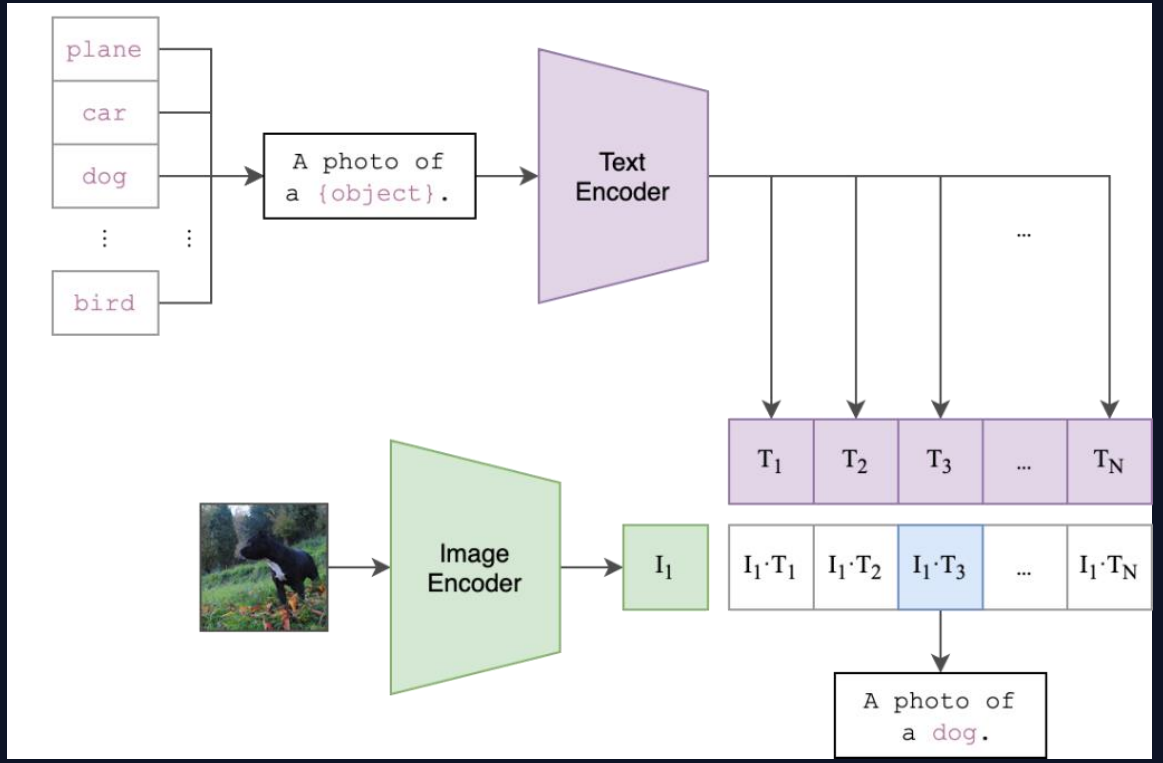


bouba

sharper

kiki

CLIP: Embeds images and text in a shared semantic space



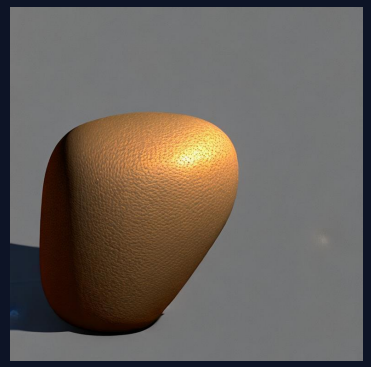
Radford et al. (ICML 2021)

SAGIVTECH
IMVC 2024

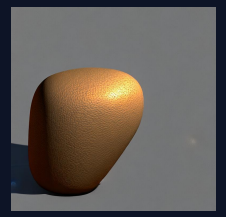
Probing VLMs for Sound Symbolism

Measuring geometric attributes of pseudowords with CLIP

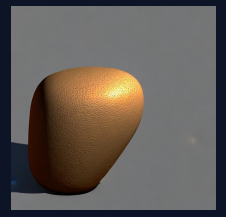
rounder



bouba

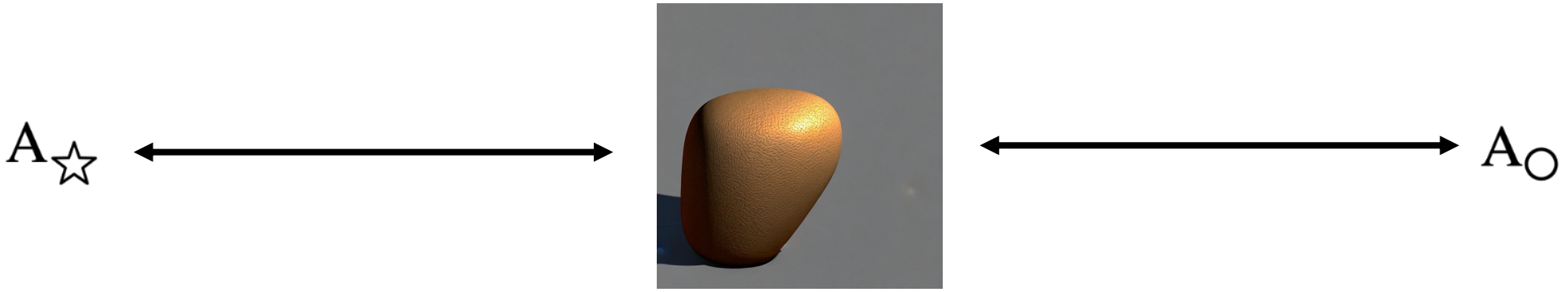


"round"



"sharp"

. . .



$A_{☆}$ = {*sharp, spiky, angular, jagged, hard, edgy, pointed, prickly, rugged, uneven*}

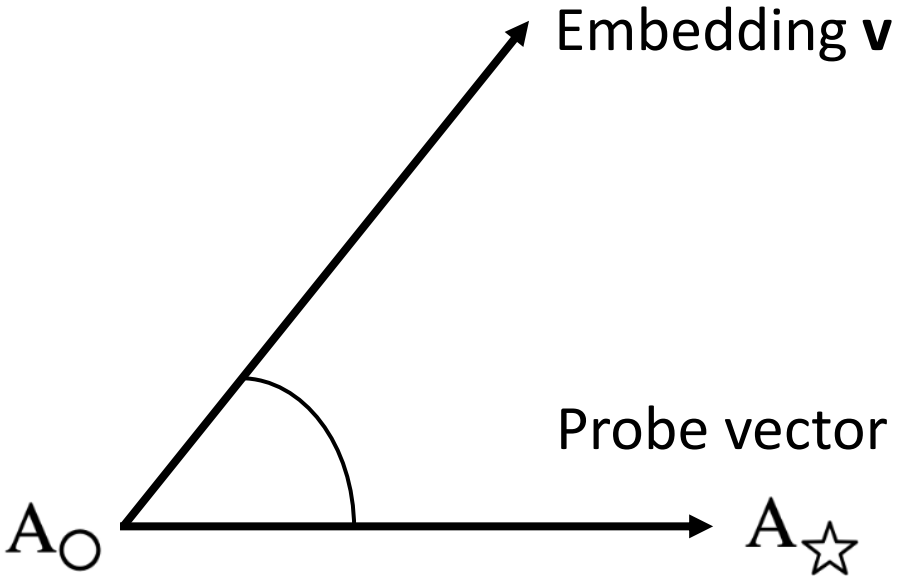
$A_{○}$ = {*round, circular, soft, fat, chubby, curved, smooth, plush, plump, rotund*}

Probe vector

$\mathrm{A}_{\bigcirc} \longrightarrow \mathrm{A}_{\star}$ $\quad w_{adj} := \sum_{\langle a \rangle \in \mathrm{A}_{\star}} \hat{w}_{\langle a \rangle} - \sum_{\langle a \rangle \in \mathrm{A}_{\bigcirc}} \hat{w}_{\langle a \rangle}$

$\mathrm{A}_{\star}$ = {*sharp, spiky, angular, jagged, hard, edgy, pointed, prickly, rugged, uneven*}

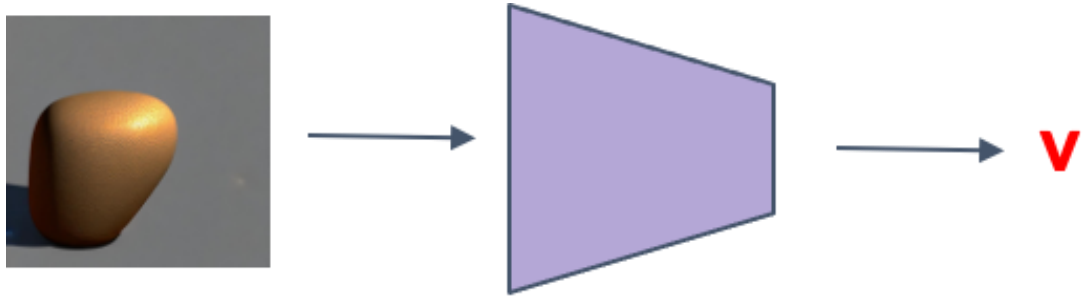
$\mathrm{A}_{\bigcirc}$ = {*round, circular, soft, fat, chubby, curved, smooth, plush, plump, rotund*}

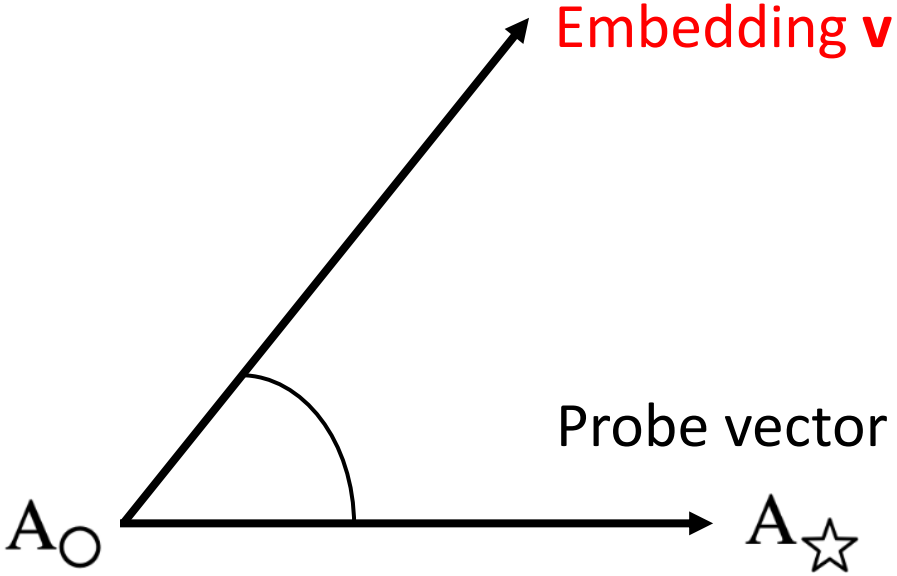


$$w_{adj} := \sum_{\langle a \rangle \in A_{\star}} \hat{w}_{\langle a \rangle} - \sum_{\langle a \rangle \in A_{\bigcirc}} \hat{w}_{\langle a \rangle}$$

$A_{\star}$ = {*sharp, spiky, angular, jagged, hard, edgy, pointed, prickly, rugged, uneven*}

$A_{\bigcirc}$ = {*round, circular, soft, fat, chubby, curved, smooth, plush, plump, rotund*}

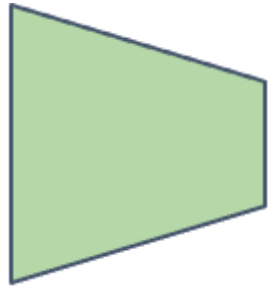


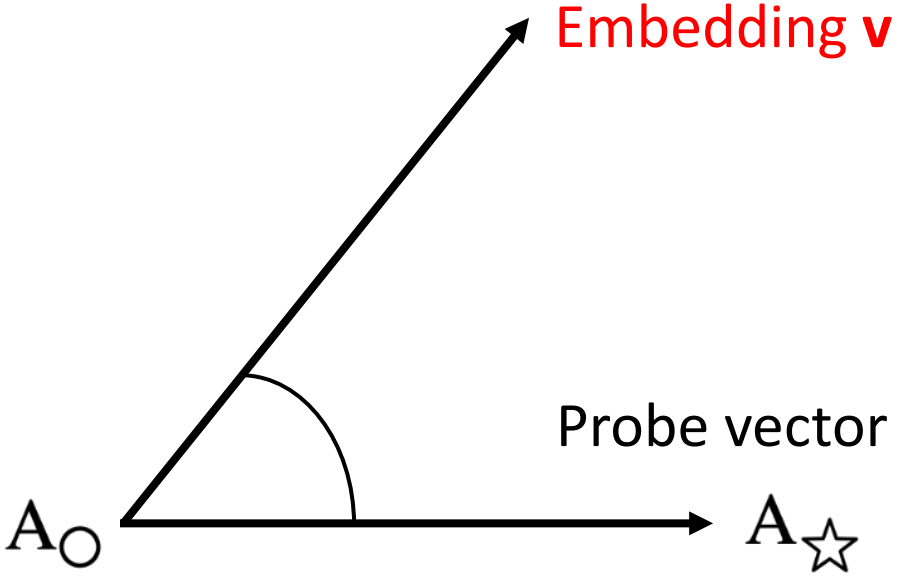


$$w_{adj} := \sum_{\langle a \rangle \in A_{\star}} \hat{w}_{\langle a \rangle} - \sum_{\langle a \rangle \in A_{\bigcirc}} \hat{w}_{\langle a \rangle}$$

$A_{\star}$ = {*sharp, spiky, angular, jagged, hard, edgy, pointed, prickly, rugged, uneven*}

$A_{\bigcirc}$ = {*round, circular, soft, fat, chubby, curved, smooth, plush, plump, rotund*}

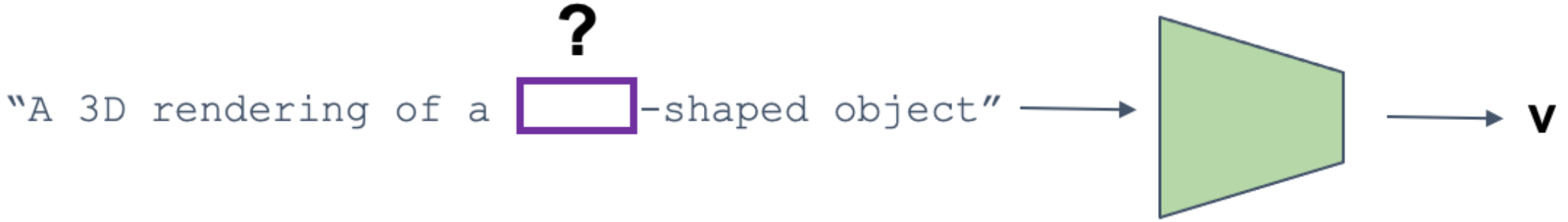
"A 3D rendering of a bouba-shaped object" →  → **v**

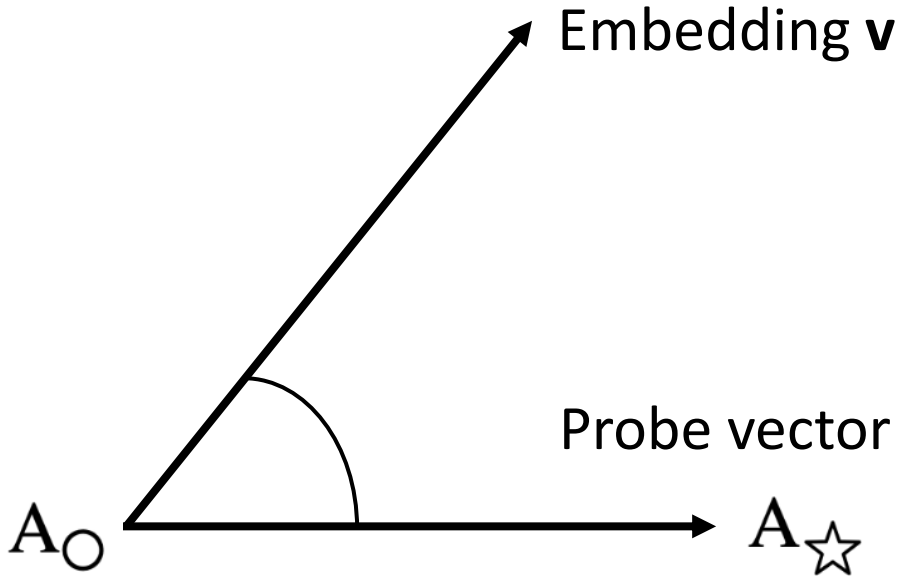


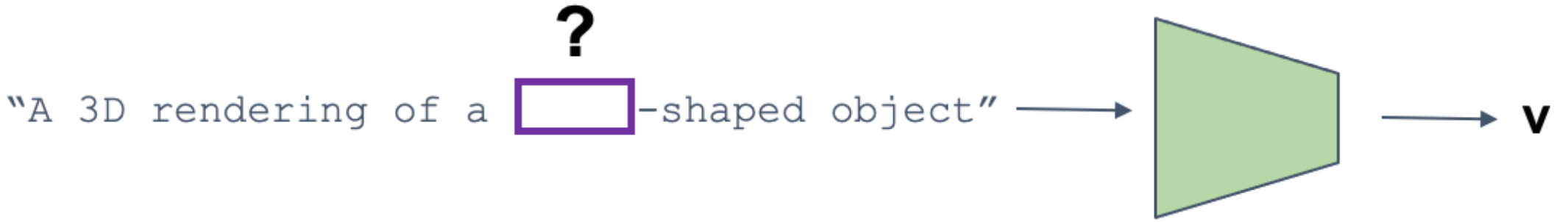
$$w_{adj} := \sum_{\langle a \rangle \in A_{\star}} \hat{w}_{\langle a \rangle} - \sum_{\langle a \rangle \in A_{\bigcirc}} \hat{w}_{\langle a \rangle}$$

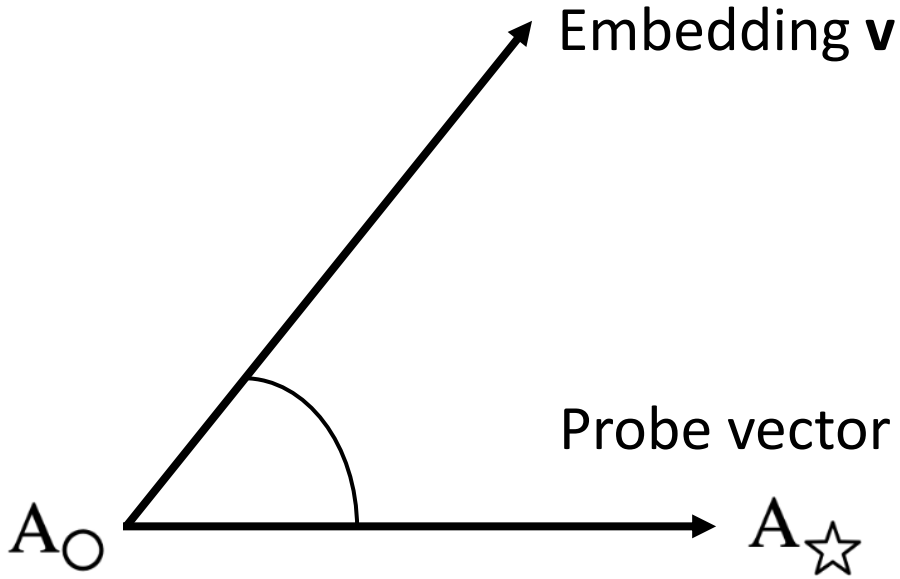
$A_{\star}$ = {*sharp, spiky, angular, jagged, hard, edgy, pointed, prickly, rugged, uneven*}

$A_{\bigcirc}$ = {*round, circular, soft, fat, chubby, curved, smooth, plush, plump, rotund*}


?
"A 3D rendering of a -shaped object"
v


Embedding v
Probe vector
A○
A☆



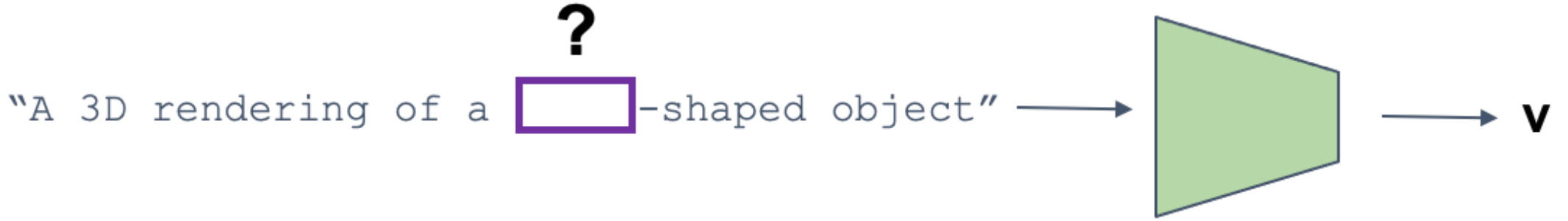


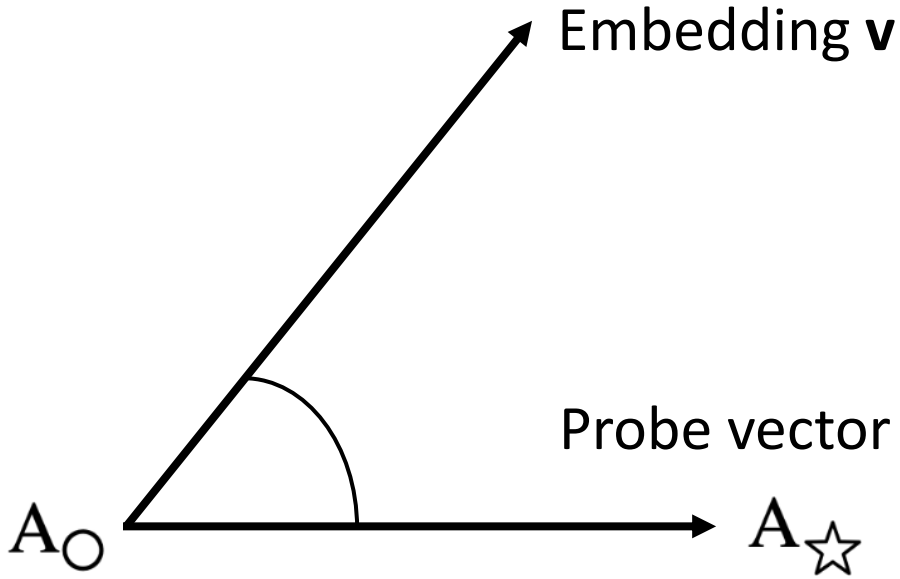
$C_{☆}$: ⟨p t k s h x⟩
$V_{☆}$: ⟨e i⟩

$C_{○}$: ⟨b d g m n l⟩
$V_{○}$: ⟨o u⟩

$V_{\_}$: ⟨a⟩

McCormick et al. (2015)





$C_{☆}$: ⟨p t k s h x⟩
$V_{☆}$: ⟨e i⟩

$C_{○}$: ⟨b d g m n l⟩
$V_{○}$: ⟨o u⟩

$V_{\_}$: ⟨a⟩

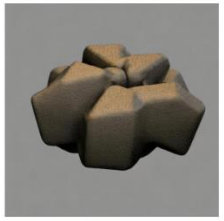
McCormick et al. (2015)

$\Psi_{☆}$:	*kitaki*	*hatiha*	*pepape*	*xisixi*	*hipehi*	*xaxaxa*	*texete*	...
$\Psi_{○}$:	*gugagu*	*bodubo*	*gunogu*	*daluda*	*momomo*	*lunulu*	*gadaga*	...

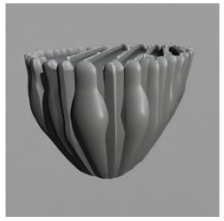
SAGIVTECH
IMVC 2024

Pseudowords sorted by geometric score

High (☆)


tepite ☆


kataka ☆

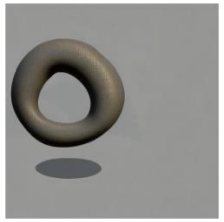

gugugu ○

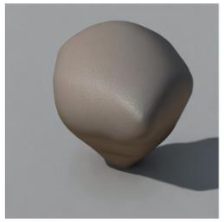

pitapi ☆

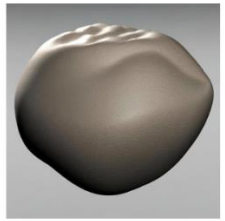

daguda ○


setese ☆

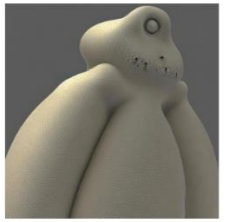
Low (◯)


nubanu ○


gomago ○


magama ○


bomabo ○

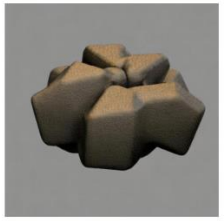

mamoma ○


lobalo ○

SAGIVTECH
IMVC 2024

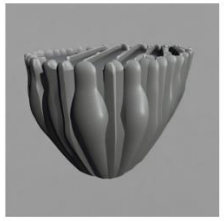
Pseudowords sorted by geometric score

High (☆)



tepite☆

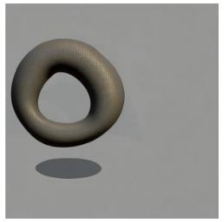
kataka☆



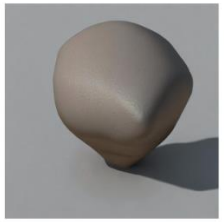
gugugu○

pitapi☆

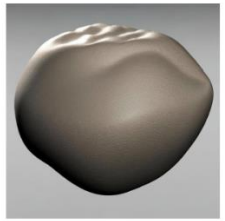
Low (◯)



nubanu○



gomago○



magama○

bomabo○

Model	AUC	τ
Stable Diffusion	0.74	0.34
CLIP	0.77	0.39
(random)	0.50	0.00

Geometric scoring

kitaki

daluda

xaxaxa

$\Psi_{☆}$:	*kitaki*	*hatiha*	*pepape*	*xisixi*	*hipehi*	*xaxaxa*	*texete*	...
$\Psi_{○}$:	*gugagu*	*bodubo*	*gunogu*	*daluda*	*momomo*	*lunulu*	*gadaga*	...

$A_{☆}$ = {*sharp, spiky, angular, jagged, hard, edgy, pointed, prickly, rugged, uneven*}

$A_{○}$ = {*round, circular, soft, fat, chubby, curved, smooth, plush, plump, rotund*}

Geometric scoring

kitaki

daluda

xaxaxa

Phonetic scoring

potato

spike

circular

$\Psi_{\star}$:	*kitaki*	*hatiha*	*pepape*	*xisixi*	*hipehi*	*xaxaxa*	*texete*	...
$\Psi_{\bigcirc}$:	*gugagu*	*bodubo*	*gunogu*	*daluda*	*momomo*	*lunulu*	*gadaga*	...

$A_{\star}$ = {*sharp, spiky, angular, jagged, hard, edgy, pointed, prickly, rugged, uneven*}

$A_{\bigcirc}$ = {*round, circular, soft, fat, chubby, curved, smooth, plush, plump, rotund*}

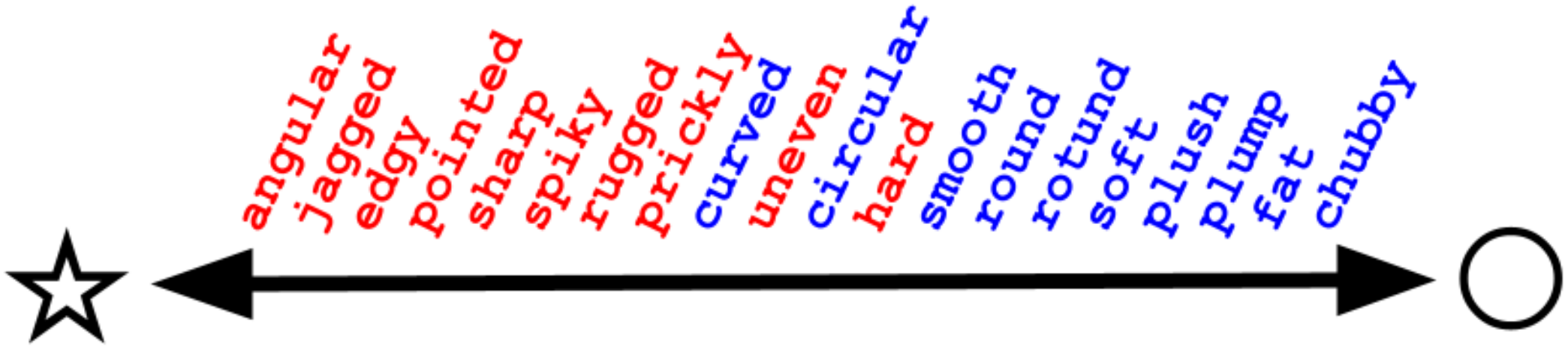
SAGIVTECH
IMVC 2024

Real English words sorted by phonetic score

	Low (◯)	High (☆)
Stable Diffusion		
Noun	*butterball, yolk, pregnancy, booger, eggnog, turnip, bellyful, crybaby, doughboy*	*shard, kite, origami, hexagon, diamond, flake, octagon, triangle, protractor, lozenge, foldout*
Adj.	*obese, chubby, stinky, pudgy, overweight, fat, pregnant, plump, drowsy, soggy, squishy*	*triangular, diagonal, angular, shattering, jagged, rectangular, edgy, housebroken, geometrical*
CLIP		
Noun	*doughboy, loudmouth, gumdrop, boogeyman, madwoman, lord, butterball, goddaughter*	*prefix, talkativeness, asker, flexibility, shears, shift, peek, slope, task, exit, hemline, tightness*
Adj.	*muggy, soggy, gloomy, grouchy, lumpy humongous, hungry, cloudless, unsmiling*	*apelike, flexible, diagonal, static, triangular, external, shipshape, interlocking, angular*

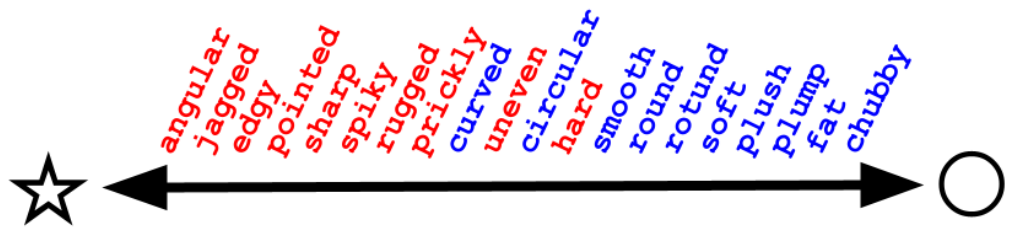
GT adjectives sorted by phonetic score

Stable Diffusion

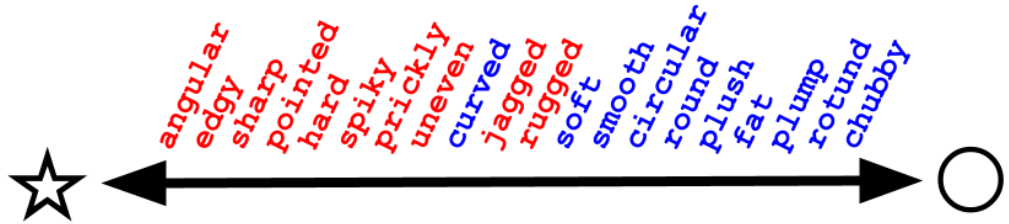


GT adjectives sorted by phonetic score

Stable Diffusion

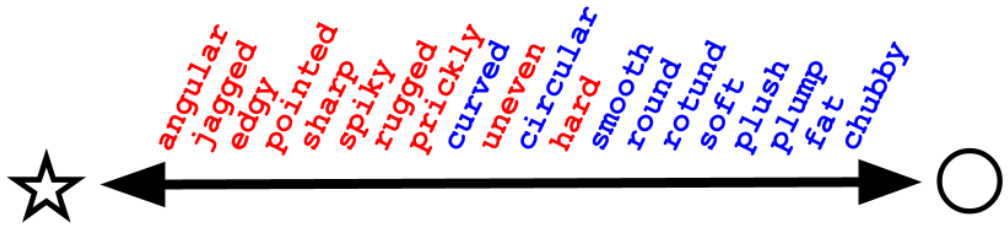


CLIP



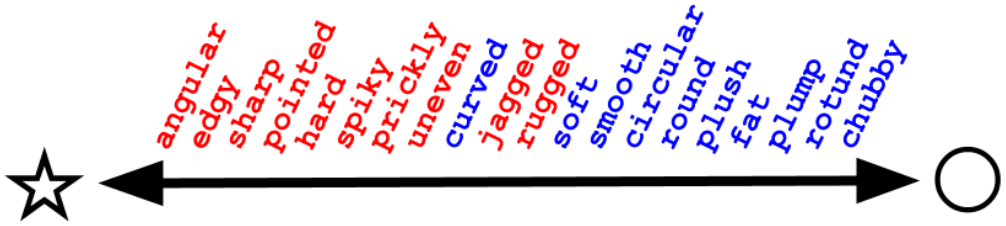
GT adjectives sorted by phonetic score

Stable Diffusion



Model	AUC	τ
Stable Diffusion	0.97	0.68
CLIP	0.98	0.70
(random)	0.50	0.00

CLIP

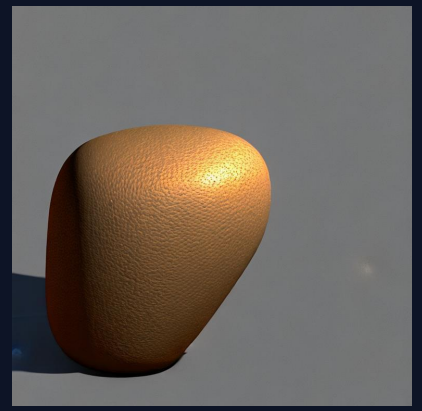


SAGIVTECH
IMVC 2024

Conclusion

We show that VLMs have learned sound symbolism, providing:

- **Insight and interpretability** for black-box models
- **A new computational tool** for studying sound symbolism in language



Questions?