

How To Build an Animated Instructor for VR Learning Applications

Sergey Kirshner
Co-Founder and CTO
Fantastic Voyage Labs

You'll Feel Spatial. We Promise.



USA • June 10-12, 2025 • Long Beach, California



Meet Celly



Why Use Avatar in Educational Applications?

- Enhanced Engagement and Motivation
 - Grab learner's attention to maintain their focus
 - Increase interaction
 - Emotional connection
- Enhanced Social Presence
 - "Persona Effect"
 - Reduce anxiety
- Personalization and Adaptivity
 - Tailored feedback
 - Adaptive guidance
- Why Implement Ourselves?
 - Wanted a non-humanoid avatar
 - Important interaction component with the user – wanted more control
 - Personal touch

Scope of the Talk: VR Animation and Basic Communication

- Creative process: developing the form and shape
- Building and optimizing a 3d model
- Voicing, lipsync, and speech recognition
- Motion
- Eye contact and gaze
- Interaction with the learner
- Future work: teaching and contextual feedback
- Not covered: hooking up avatar to LLMs or knowledge bases

Acknowledgements



Taras Voznyi conceived and created Celly's visuals and models



Pavel Koutcherina built Celly's interaction, movement, and behavioral components

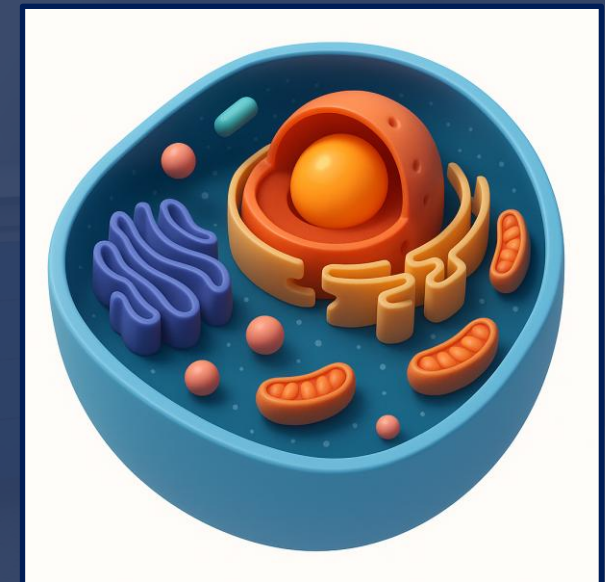
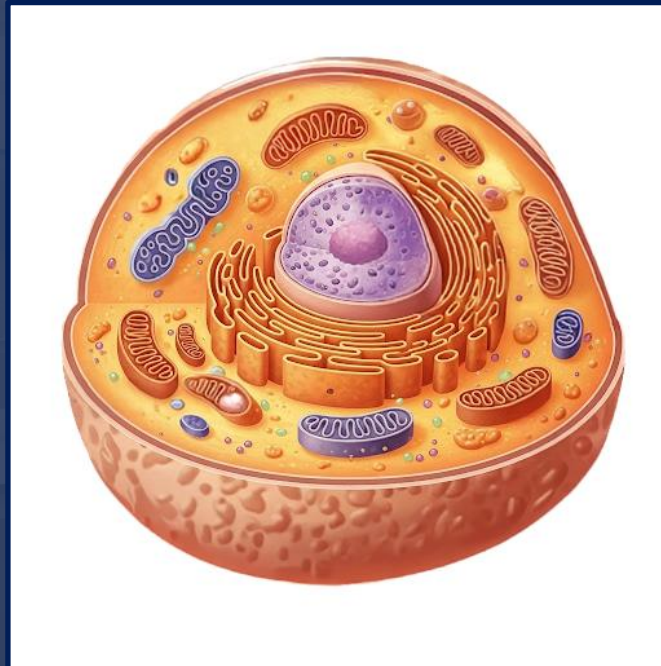
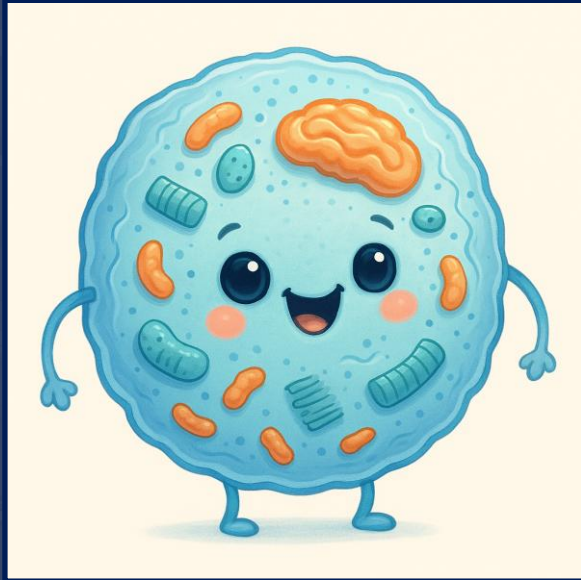


Daniel Yu, Connor Gore-Rodney, and Sergiy Putiak contributed to Celly's creation and improvements to her behavior

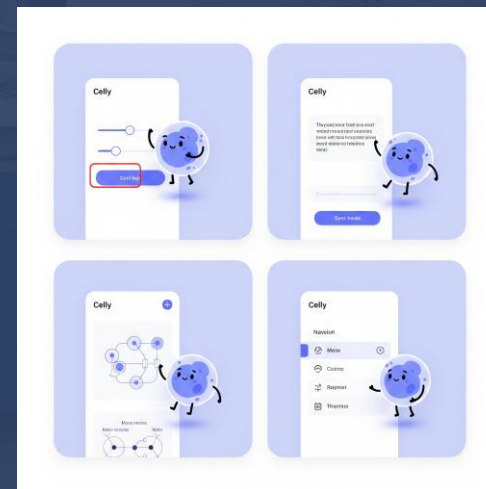
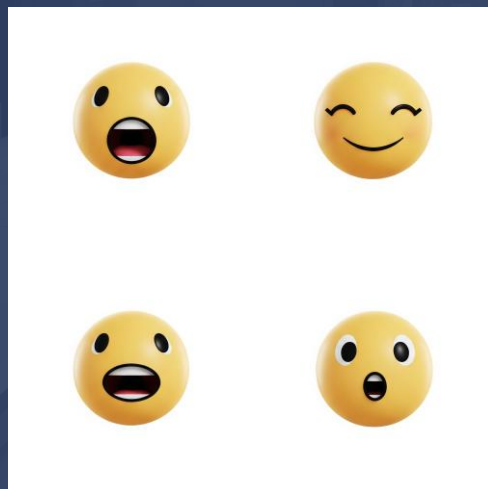
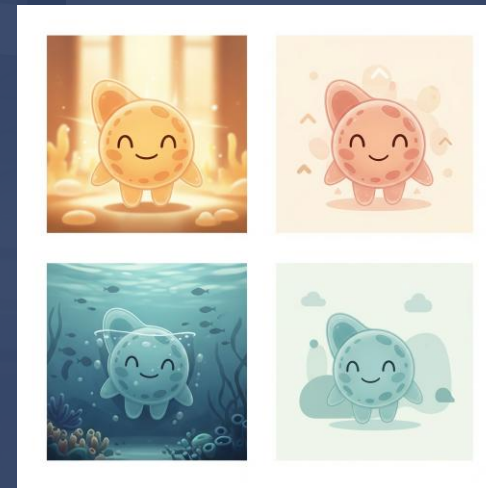
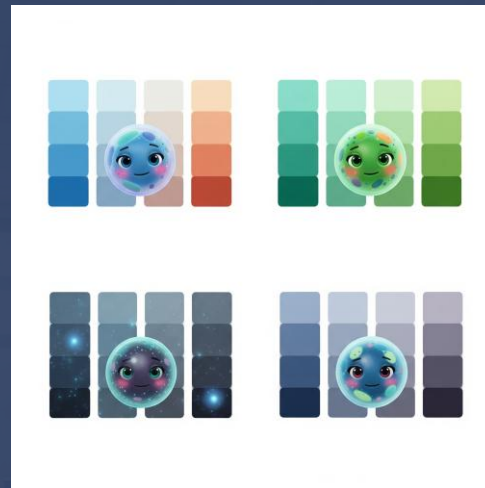
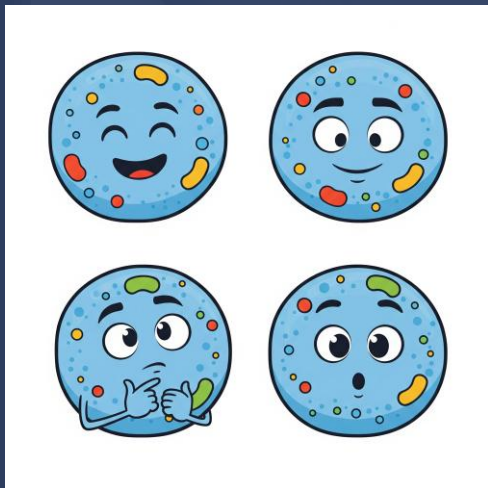
Developing the Form

- **Goal**
 - To enhance user engagement, making learning more interesting, effective, and comfortable, make it easier for the user to absorb information and gain skills.
- **Design requirements**
 - **Core Concept:** design must be based on a biological cell
 - **Appearance:** friendly, trustworthy, non-distracting
 - **Functionality:** ability to speak, fly/move freely in space
 - **Impact:** creates a comfortable atmosphere where users aren't afraid to make mistakes and are encouraged to listen to its advice

Start With Reference Images



Create Reference Mood Board



(Unsuccessful) Attempts with GenAI



Finding the Perfect Look: Iterative Process

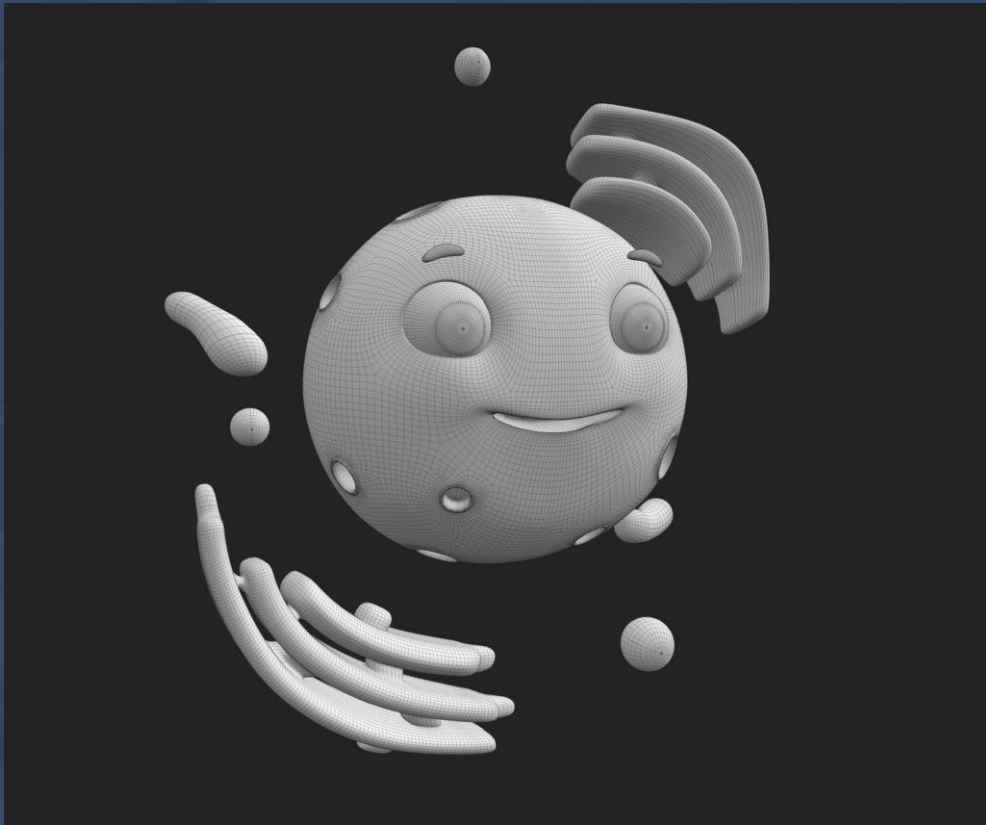


Creating and Optimizing 3D Model

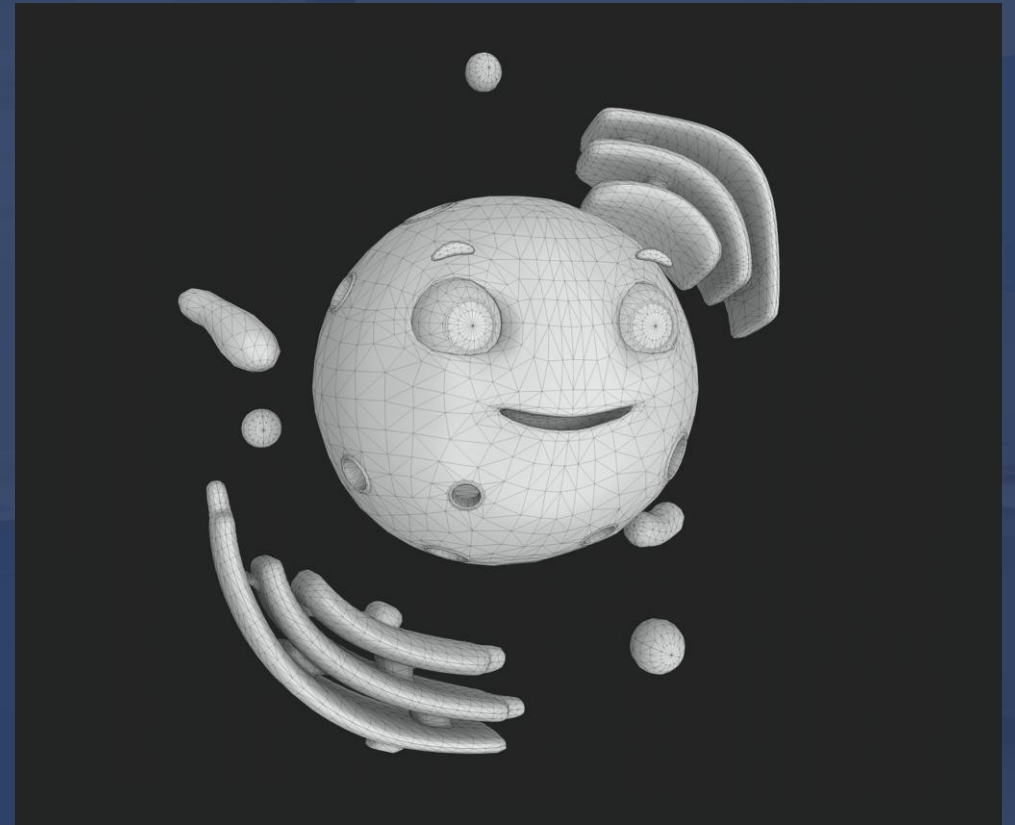
- 2D -> 3D (sculpting): build a high poly 3D model
- UV Unwrapping to extract 2D surfaces
- Baking: simplify the differences between the high poly and low poly 3D models
- Texturing: add colors, roughness, albedo etc to surfaces

3D Modeling

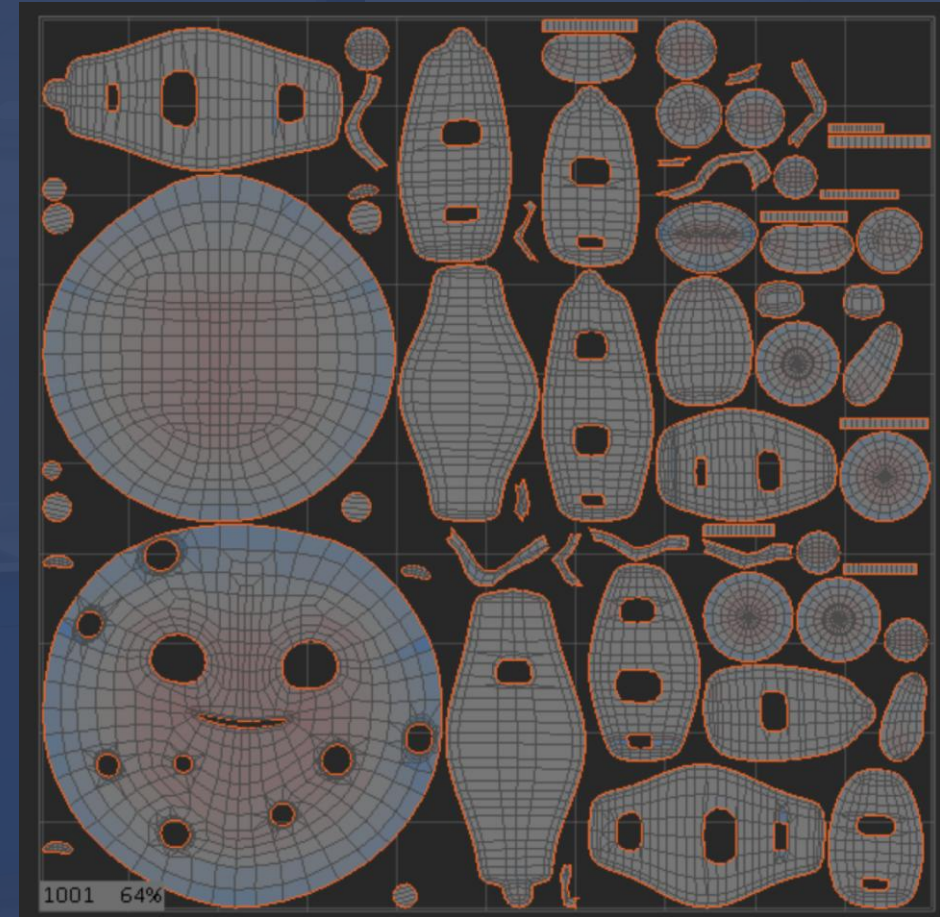
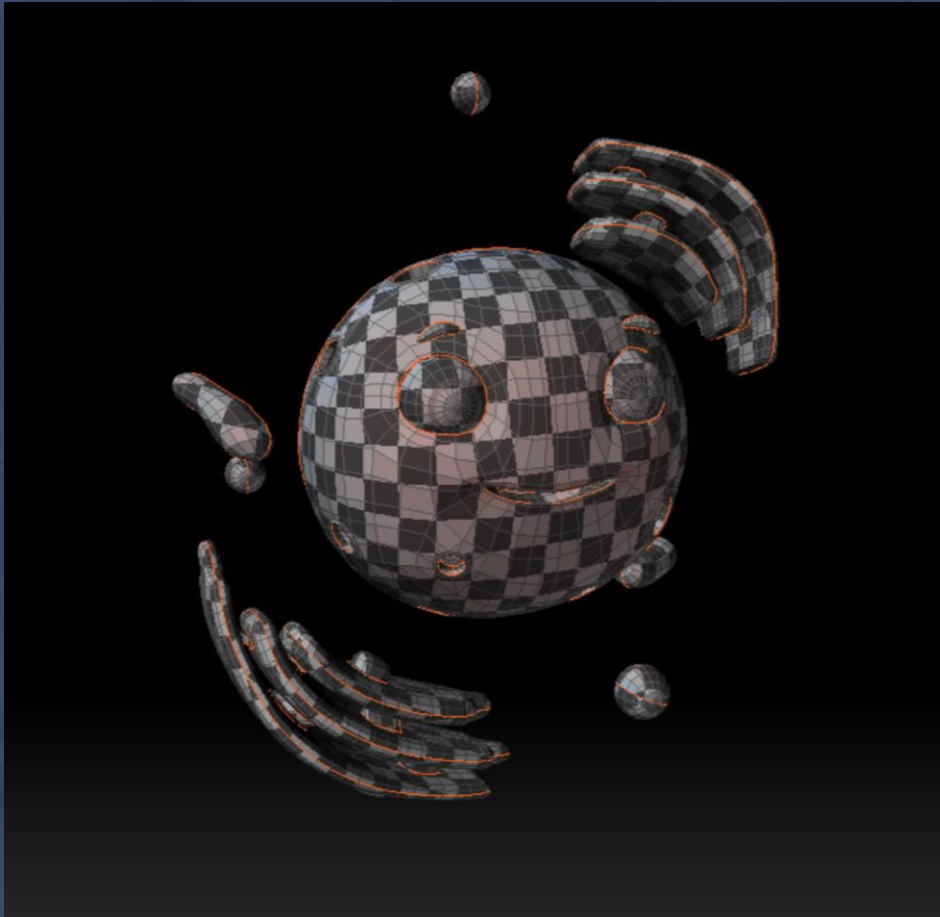
High Poly model



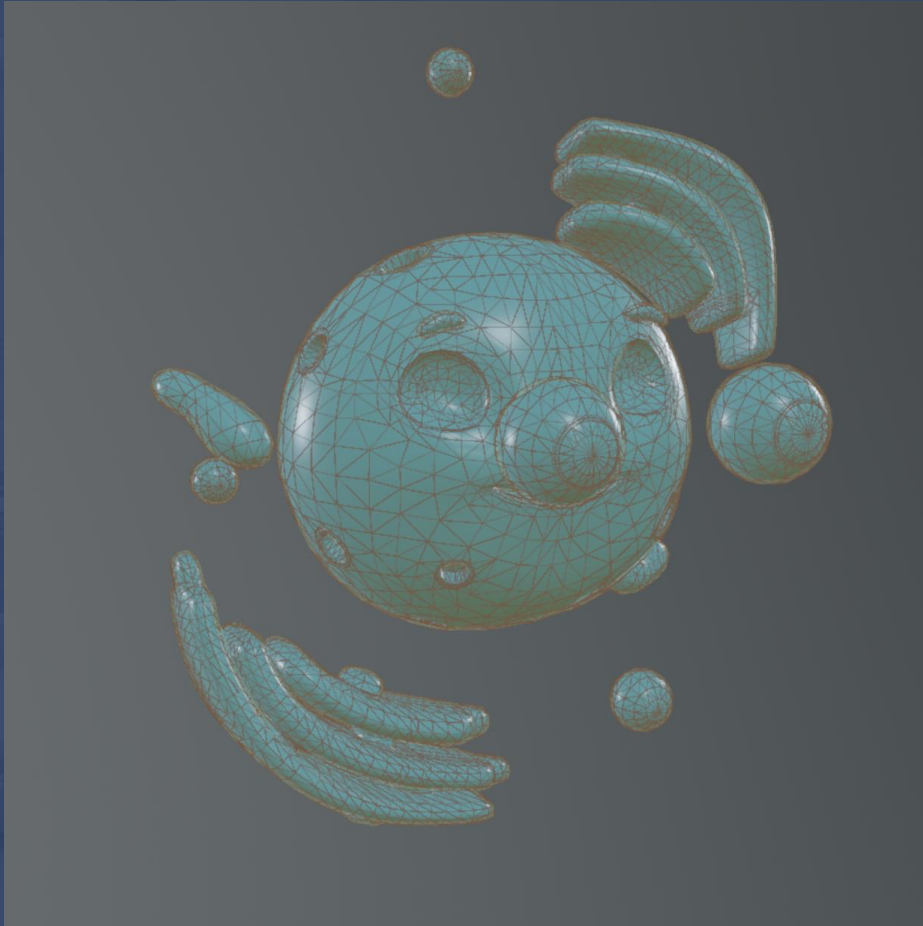
Low Poly model



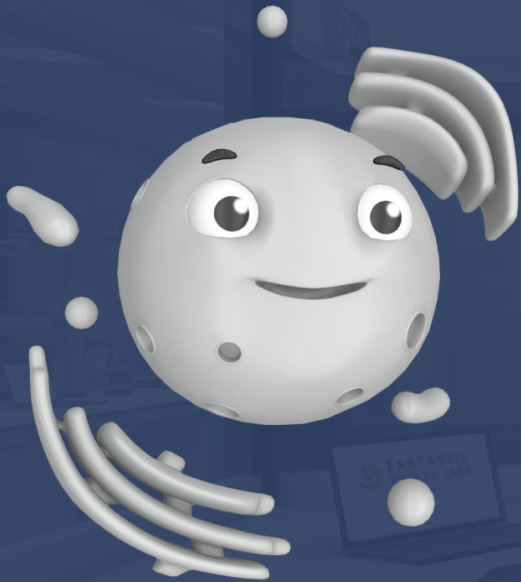
UV Unwrapping



Baking



Texturing



Making Celly Talk: Voicing

- Speech synthesis
 - Off-the-shelf text-to-speech (TTS) tools – wit.ai, free, easy integration with Unity
 - Library of voices requires some experimentation
- Cloud vs Local?
 - Pre-baked narration for static feedback
 - No dependence on internet connection
 - Requires building pipelines to generate and update static narrations
 - Only works with static text
 - Cloud for experimentation and TTS from dynamic text
 - Can also use a TTS solution on the device
 - Space and cycles

Synchronizing Speech with Face



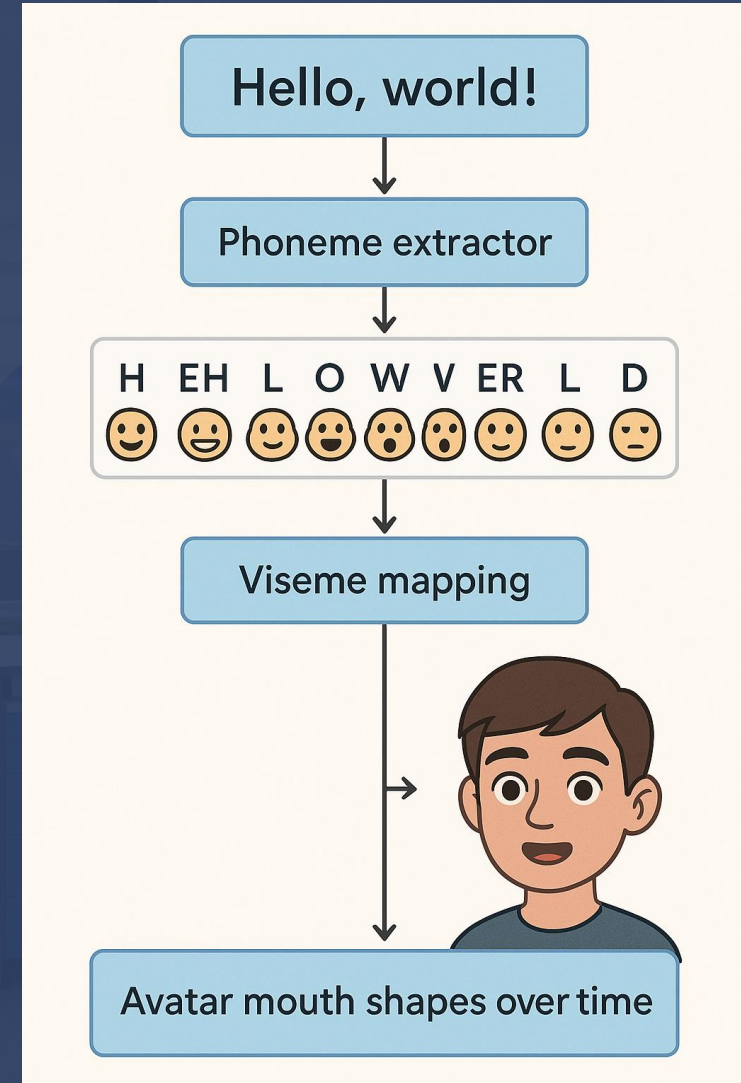
LipSync ON



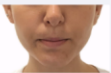
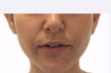
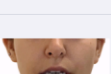
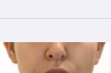
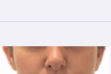
LipSync OFF

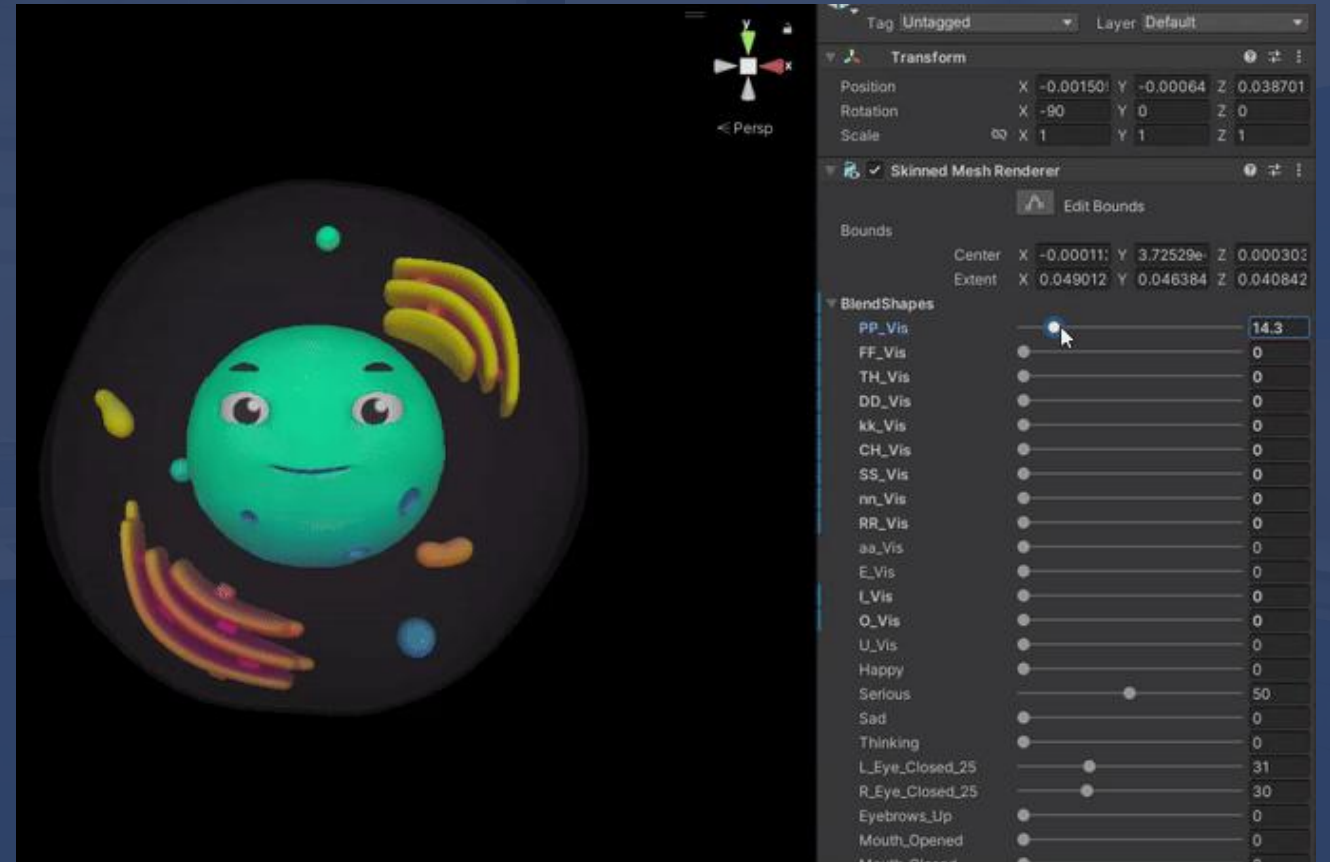
Making Celly Talk

- Phrases are broken down into phonemes
 - 44 phonemes in English
- Phonemes are mapped into transitions between visemes
 - Mouth or face shapes
 - Only 14 visemes + neutral
 - Need to create the corresponding blendshapes for the avatar
- Using uLipSync open source framework [\[github.com/hecomi/uLipSync\]](https://github.com/hecomi/uLipSync)



Blendshapes

Viseme Name	Phonemes	Examples	Mild Production
sil	neutral	(none - silence)	
PP	p, b, m	put, bat, mat	
FF	f, v	fat, vat	
TH	th	think, that	
DD	t, d	tip, doll	
kk	k, g	call, gas	
CH	tS, dZ, S	chair, join, she	
SS	s, z	sir, zeal	



Making Celly Listen

- Local (offline) model for speech recognition
 - Recognissimo [\[Unity Asset Store\]](#)
 - Does not rely on the internet connection
 - Does not want user utterances sent to a 3P provider
- Wake word to get Celly's attention
 - Constantly listening for wake word/phrase (Hey, Celly!)
 - 10 second buffer for commands/interaction once activated
- During the attention window
 - Map transcribed text to commands or actions
 - Allow chaining of the commands

Making Celly More Real

- Gaze
 - Blinking, pupil twitching
 - Following learner's gaze
 - Looking down if too close to the learner
 - Realistic Eye Movement [[Unity Asset Store](#)]
- Motion
 - Idle gyrations
 - Figure eight pattern
 - Subtle organelle movements



Making Celly Interact

- Celly can understand commands
 - Permits navigation through Celly: teleporting, turning, opening menus
- Celly can guide the learner through the scenario
 - Narrates the scenario instructions
 - Alerts the learner of errors

Making Celly Interact



Making Celly Teach

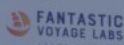
- Contextual hints based on the progress in the scenarios
 - Can also point
- Responses based on learner's gaze
 - Responses based on the context of the room, e.g., locations and proximities of the objects and highlighting
- Answer questions relevant to the scenario
- Contextual feedback
 - e.g., slow down
- Be multilingual
- Add fun to learning! Emotions, funny movements, phrases

Summary and Future Directions

- Cannot have learner listen to the void – difficult to stay focused
- Relatively straightforward to add your own avatar to tailor it to your application
- Iterative process
- Small touches are needed for realism
 - Lipsync
 - Realistic eye movements



the Future of
Skills



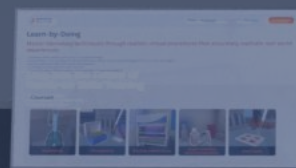
Virtual Lab, Real Skills

Virtual training to master hands-on laboratory skills featuring a lifelike laboratory, personalized instruction, and adaptive content

BUSINESS TEMPLATE
HEADLINE



FANTASTIC VOYAGE LABS



FANTASTIC
VOYAGE LABS
LEARN BY DOING



Meet Celly at Table S4