**1) Team member names**

Ying Li, Aman Kaleem

**2) Project name (if you have one) and very short IDEP 'title';**

*Smrti | Yi* - “Explore a new way to reconstruct and present memories”

**3) A paragraph or two (bullet points are fine) describing the problem you've identified;**

The encoding of memory is intricate, involving the vibrant but malleable sensory details of our experiences. Context is a cornerstone in how memories are embedded and later accessed, with the mingling of similar contexts occasionally leading to recall mistakes. Emotionally charged memories are often the most poignant, yet they carry the risk of coloring our recollections. Presently, we encapsulate memories through written words, photographs, and videos. Venturing beyond, we are exploring the next phase of memory capture: the reconstruction of memories. This emerging technology, while versatile like video, is primarily focused on recreating experiences as rich and personal as the pages of a photo album.Our goal transcends mere replication of traditional memory preservation methods; we strive to infuse emotions and sensory details into our reconstructions,

3) **A paragraph or two on your solution/prototype;**

Our solution is at the nexus of cognitive science, artificial intelligence, and brain-computer interface technology. Leveraging the capabilities of machine learning and generative AI, this project transforms auditory inputs into visual representations of memories with real-time fluidity. It's a system engineered to decode and visualize auditory and non-verbal cues, including emotional undertones and subtle body movements, and reconstructing them into visual narratives that reflect the intricacies of personal recollections. This pursuit aims to investigate the body of memory. By piecing together the sensorial fragments of memory, it endeavors to weave complex individual recollections into visual narratives. The correlations between text and image are learned by the AI models in a high-dimensional latent space, through interpreting tens of millions data pairs.  After an auditory input we deconstruct memories into their fundamental elements: characters, scenes, and sequences of events. Each segment is carefully rebuilt and then skillfully merged using video production techniques. In the reconstruction process, values are assigned to aspects such as non-verbal cues, emotional resonances, and the subtleties of body language, ensuring these elements play a crucial role in the fidelity and richness of the recreated memory.

**4) Specific challenges you foresee and questions you have for reviewers.**

* What can be some critical factors to consider in addition to what we present, when it comes to memory reconstruction?
* What approaches can be developed to navigate ambiguities and accurately reconstruct memories despite potential inaccuracies in the user's recollection?
* How can we ensure the privacy and security of the data involved in the memory reconstruction process, and what ethical considerations must be addressed when using such deeply personal data?.

**Smrity-Yi**

**The Ontology of Collective Memory**

**Product/service or company/NGO:**

We explore a new way to reconstruct and present memories. Our goal transcends mere replication of traditional memory preservation methods; we strive to infuse emotions and sensory details into our reconstructions.

The encoding of memory is intricate, involving the vibrant but malleable sensory details of our experiences. Context is a cornerstone in how memories are embedded and later accessed, with the mingling of similar contexts occasionally leading to recall mistakes. Emotionally charged memories are often the most poignant, yet they carry the risk of coloring our recollections. Presently, we encapsulate memories through written words, photographs, and videos. Venturing beyond, we are exploring the next phase of memory capture: the reconstruction of memories. This emerging technology, while versatile like video, is primarily focused on recreating experiences as rich and personal as the pages of a photo album

**Type of project & funder:**

We can target several types of funders, each with different motivations and interests in the project. Here’s a breakdown of potential audiences  and funders

1. **Venture Capitalists (VCs) Specializing in Tech Innovations**
   * **Focus**: Venture capitalists such as Andreessen Horowitz, Sequoia Capital, and Kleiner Perkins, who are known for their investments in breakthrough technologies, are particularly suited as funders. These VCs are typically attracted to ventures that promise substantial returns through disruptive innovations. The unique aspect of this memory reconstruction technology, which transcends traditional memory preservation and introduces a more dynamic and emotionally rich user experience, positions it as a high-potential investment.
2. **Technology Corporations with Strategic Investment Arms**
   * **Focus**: Large tech corporations like Google and Microsoft, which operate venture arms like Google Ventures and M12, respectively, could find strategic value in this technology. These companies often invest in startups that align with or enhance their existing technological ecosystem. Given that this project integrates advancements in AI, AR, and VR to revolutionize memory preservation, it aligns well with their ongoing innovations in these domains. The potential for integration into healthcare technology further broadens its appeal, offering synergistic opportunities in developing new therapeutic tools based on memory reconstruction.
3. **Industry Innovators in AI, VR, and AR Platforms**
   * **Focus**: Companies and platforms specializing in artificial intelligence, virtual reality, and augmented reality technology stand as significant potential funders and collaborators. The project's reliance on cutting-edge AI to encode and recreate personal memories offers a novel application area for VR and AR technologies, opening avenues for immersive memory experiences that can transform how personal histories are interacted with and understood.

**Opportunity statement:**

Our project resides at the dynamic intersection of cognitive science, artificial intelligence, and brain-computer interface technology. It innovatively transforms auditory inputs into visual representations of memories with unparalleled real-time fluidity, using the latest advancements in machine learning and generative AI. This system is meticulously engineered to decode auditory signals and instantaneously visualize them as vivid, living memories.

Central to our initiative is the exploration of memory as a bodily phenomenon—continuously shaped by our interactions with our surroundings and the technologies we engage with. This project pushes beyond traditional memory aids, seeking to capture and reconstruct the forgotten fragments of everyday life into visual narratives that address the common amnesia of our daily experiences.

We adopt a concept we term the 'bio-politics of collective memory,' where technology transcends its role as a mere interface and becomes a fundamental part of communal life. This approach challenges the prevailing tech paradigms that often lead to isolation, promoting instead a design philosophy that fosters connectivity and shared experiences. This radical rethinking aims to shift the role of technology from enhancing individual productivity to enriching the communal fabric of society.

**Impact and Applications:**

* Therapeutic Use: Our technology acts as a potent tool for emotional healing, enabling individuals to revisit and process past traumas in a secure environment.
* Educational Advancement: It serves as a novel resource for delving into human cognition, the learning process, and the emotional effects of memory, thereby enhancing educational methodologies.
* Cultural and Creative Expression: The technology provides creatives and filmmakers with new storytelling tools, making it possible to create documentaries and narratives that revitalize historical events and personal stories, thus deepening cultural insights and perspectives.

By leveraging this advanced technology, our project not only promises a future where experiences are preserved but also where they can be vividly revisited and felt, thereby deepening our collective understanding of memory's role in shaping human experience.

**Why Speculative Design**

Speculative design serves as a transformative approach within the broader design field, emphasizing the power of imagination and critical thinking to explore and question potential futures. Unlike traditional design, which often concentrates on solving specific, immediate problems or enhancing consumer products, speculative design offers a more provocative role. It challenges established norms and conventions, using design as a critical tool to interrogate the implications of new technologies and societal trends, and to envision alternative possibilities. This approach does not just seek to meet the existing needs of society but to illuminate and question underlying assumptions about those needs and the technologies developed to meet them.

Our venture into the realm of speculative design was motivated by a desire to address the increasingly prevalent phenomenon of isolation in a connected world—where technology often leads to people being "alone together." We aimed to rethink the role of technology in our lives, focusing on designs that engage the human body and mind at their core, fostering genuine connectivity and community. Entering this space allowed us to explore uncharted territories where the rules are not yet defined, providing us the freedom to experiment with ideas that challenge how technology interacts with and shapes our communal and individual experiences. This approach is not only about creating new products but also about crafting new paradigms for how technology can enhance, rather than diminish, our human experience.A purple cone with white lines around it

Description automatically generated

Memory is a deeply human construct intricately linked to the fabric of our identities. It shapes not just who we are but how we connect with others and our surroundings. Given its profound impact on identity, any project dealing with memory must consider implications that extend far beyond the conventional metrics of market value and capital gain. In this context, our approach to the memory reconstruction project adopted a perspective that values the human experience above financial outcomes. This approach recognizes that the true worth of memory-enhancing technologies lies in their capacity to enrich human lives, foster deeper understanding among people, and bridge the gaps between generations and cultures. Our project treats memory not merely as a source of data to be mined but as a sacred space that commands respect and sensitivity. Therefore, we have crafted a preferable future within the realm of speculative design. This approach allows us to explore and shape how memory technology should be ethically and thoughtfully integrated into society, ensuring that it enhances rather than exploits our human experience.

**Solution/prototype**

Our solution is at the nexus of cognitive science, artificial intelligence, and brain-computer interface technology. Leveraging the capabilities of machine learning and generative AI, this project transforms auditory inputs into visual representations of memories with real-time fluidity. It's a system engineered to decode and visualize auditory and non-verbal cues, including emotional undertones and subtle body movements, and reconstructing them into visual narratives that reflect the intricacies of personal recollections. This pursuit aims to investigate the body of memory. By piecing together the sensorial fragments of memory, it endeavors to weave complex individual recollections into visual narratives. The correlations between text and image are learned by the AI models in a high-dimensional latent space, through interpreting tens of millions data pairs.  After an auditory input we deconstruct memories into their fundamental elements: characters, scenes, and sequences of events. Each segment is carefully rebuilt and then skillfully merged using video production techniques. In the reconstruction process, values are assigned to aspects such as non-verbal cues, emotional resonances, and the subtleties of body language, ensuring these elements play a crucial role in the fidelity and richness of the recreated memory.

Our biggest contribution is to integrate a mixture of technological solutions to create a proof of experience

The bottom layer is a collection of fundamental models that can generate high-quality contents from the research and open source community.

The middle layer is fine tuning which aims at increasing consistency of character and scene generation based on the story in context. On the top of the stack is to combine traditional film production techniques like video editing and cinematography techniques with generative models to make the experience more believable.

A screenshot of a computer

Description automatically generated

Figure: Tech Stack

A computer screen shot of a flowchart

Description automatically generated

Figure: Image Construction Flow

A diagram of a person

Description automatically generated

Figure: Overview of the workflow for proof of experience

To produce a memory experience, it starts with collecting verbal descriptions of a piece of memory. It is a two-way conversation, conducted in a way like sharing with a friend what you remember, the listener generally would follow up with questions for more details. We found being descriptive is key to increase text fidelity and the overall quality of visuals generated

The audio then is transformed to text format and goes through script analysis.

To increase immersiveness and faithfulness, we use a reference engine to search image archives, real protagonist photos and even detailed description of environmental characteristics (e.g what moss looks like). The reference data would be used as image generation guidance.  After an image is generated, sequences of images would be extrapolated to videos and stitched together. There is quite a lot of subjective choice made right now for our proof of experience. In the future, we will seek alternative solutions for users to generate them.

Lastly, the video goes through post processing to add emotions based on the semantics.  Then video will be synced with the audio voice to complete the story generation.

In summary it is an iterative decomposition and recomposition process.

Limitations

The current approach has many touch points where we make choices for the protagonist. This can make high-quality outcomes but in the long term, we have to give the control back to the users.

Current generative models can potentially generate outcomes that break laws of physics. This is an open research topic as the models are only trained from images not how the world operates. As a result, from time to time, generated images might show implausible contents. We usually hide such images from users. In the long term, we need a solution here by putting guardrails before showing images to end users.

Testing/validation:

We started with testing the reconstruction on ourselves, then expanded it to family and friends.

Through testing on ourselves, we reached important learnings quickly:

1. We were usually surprised at first what is being generated based on our memories. It really depends on what the focus of mind was when the event happened and how descriptive the protagonist is narrating the events and environment. So a key process needs to allow the protagonist to iterate and refine the narrative. But on the other hand, it would increase the chance of alternating what one really saw. But this is a personal choice that we should not intervene. Because through this process, one can usually discover rich information about themselves, how the mind works and where the heart was. And this is exactly the point of reconstructing memories.
2. It is very challenging to reconstruct faces for someone we are familiar with. The familiarity raises the quality bar and it is uncomfortable to see faces that are not faithful to how they really looked. In addition, as we age, how we remember their faces will change as well, subjected to many cognitive biases. For example, we remember recent things better but at the same time, we put more weight on memories we would like to remember, so called rosy retrospection bias. In my case, I would rather see my parents when they were younger than now, when they would pick me up from high school. So I am never satisfied with the reconstructed faces of my parents. So the decision we made is to deemphasize facial faithfulness and immature reconstruction details in the project, rather we would emphasize the vibe and visual memory triggers.

Through testing on family and Friends, we learned:

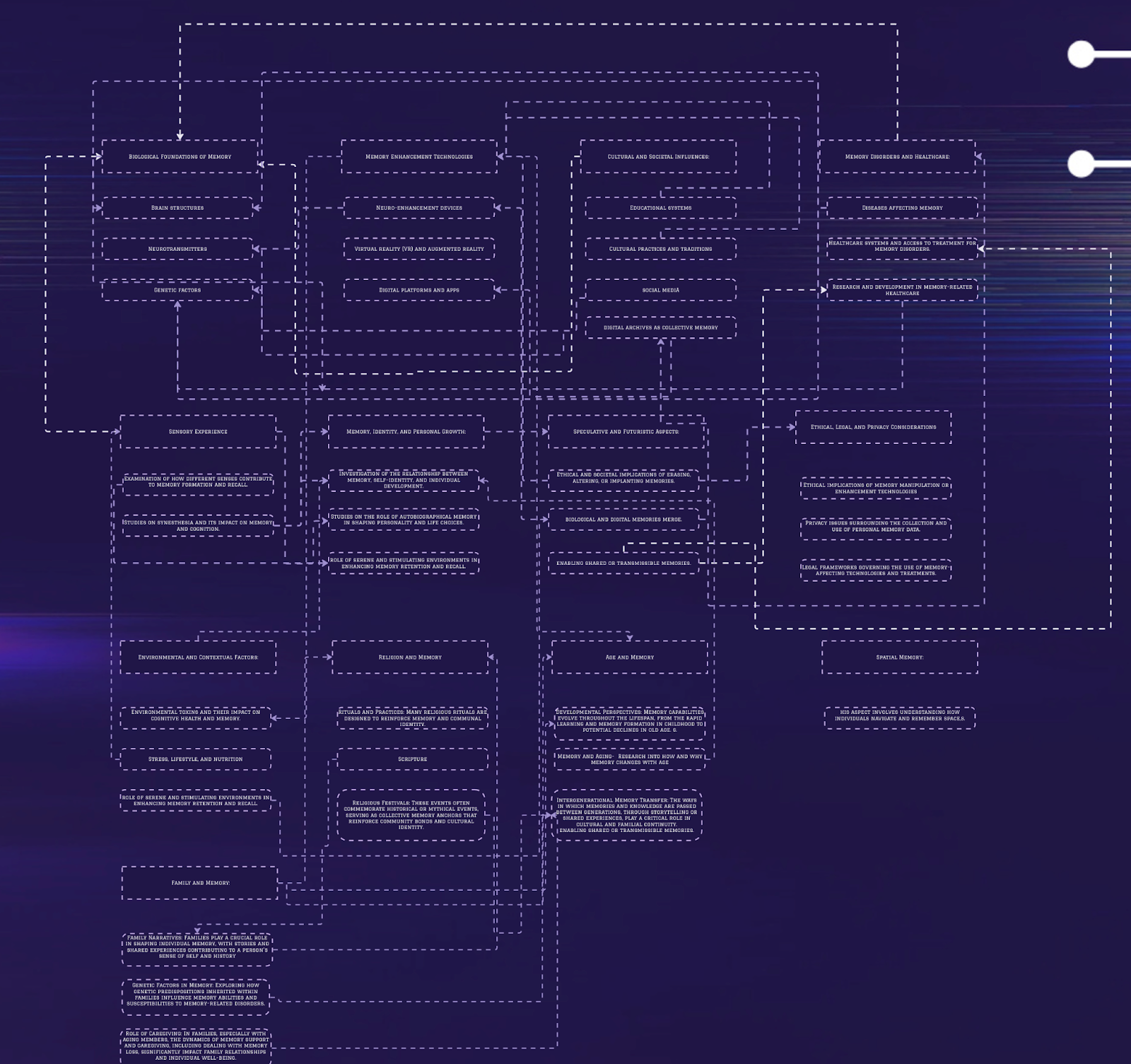
1. People often remember events through their lens, which can be shaped by factors such as personality, emotions, and individuality. For example, my best friends and I spent an extensive period of time together. But when we talk about our impressions for each other, we remember different moments. Thus reconstructing memories and sharing with each other can usually surprise us in different ways.
2. The generation is through probabilistic models. Depending on the volume of data the model has seen, there can be data bias. For example, the mainstream models are good at generating scenes and characters in the context of western countries, but do a poor job of generating context from India and China. It may not see much of recent images from those countries thus generated images are biased towards poor communities in the 90s.

**System map, if helpful, reflecting different scales:**

The potential for collective memory networks, enabling shared or transmissible memories

The evolution of memory where biological and digital memories merge.

Ethical and societal implications of erasing, altering, or implanting memories.



Design research

Experts consulted

**Harvard University**

* **Daniel Schacter** - Professor specializing in memory research, Department of Psychology
* **Arthur Kleinman** - Professor of Medical Anthropology, Psychiatry, and Global Health and Social Medicine
* **Michael D. Smith** - Professor of Engineering and Applied Sciences
* **Todd Zickler** - Director of the Harvard Computer Vision Laboratory

**Harvard Medical School**

* **Mark Dubbelman** - Postdoctoral Researcher at Mass General Brigham

**Harvard Business School**

* **Vinay Mehra** - Coordinator, Future of Media Lab

**MIT Media Lab**

* **Gabriela Bila Advincula** - Research Assistant

**Zhejiang University**

* **Yiyi Liao** - Assistant Professor, leads the X-Dimensional Representations Lab (X-D Lab)

**Huawei - Noah’s Ark Lab**

* **Hao Chen** - Computer Vision Researcher

**Literature: Reconstruction & Generation Models**

**Understanding Memory:**

The intricate process of memory encoding is fundamental to how we perceive, process, and retain information, forming the backbone of human cognition and learning. This overview categorizes memory encoding into distinct types, each tailored to handling various forms of data, from visual and auditory stimuli to complex semantic and emotional information

A screenshot of a computer program

Description automatically generated

1. **Visual Encoding** focuses on the conversion of visual input into memorable data. It is greatly affected by how attentively we process visual information, the depth of such processing, our emotional state at the time of encoding, and neurotransmitter activities. These factors contribute to how vividly and for how long these visual memories are preserved in our brain.
2. **Elaborative Encoding** involves connecting new information to already existing knowledge, enhancing recall by embedding new data into a network of previously understood concepts. This type of encoding leverages existing knowledge, cognitive associations, and the level of attention given to the new information, facilitating more profound processing and better integration of the new with the old.
3. **Semantic Encoding** deals with encoding information by its meaning rather than just the sensory input. It allows for the information to be associated with existing knowledge and contexts, making use of strategies like chunking and mnemonics. The meaningfulness of the information, its relation to what we already know, and the cognitive strategies employed are critical in how effectively semantic memories are formed.
4. **Acoustic Encoding** is about the encoding of sounds, particularly words and other auditory information. The factors influencing this type of encoding include auditory processing abilities, phonological awareness, the level of attention, and the repetition of the sound. These elements help in forming lasting auditory memories, essential for tasks like learning new languages or remembering names.

Together, these categories illustrate a comprehensive framework of how memory works in humans, highlighting the complex interplay between sensory processing, cognitive operations, and emotional factors. This understanding is crucial not only in cognitive psychology but also in designing educational and therapeutic interventions that enhance memory retention and recall.

**Seven Sins of memory**

The concept of the "Seven Sins of Memory" was introduced by psychologist Daniel Schacter and highlights the ways in which our memories can fail us. Prof. Daniel Schacter  was part of the advisory committee and has been present in our reviews.

Seven Sins are not just mere errors but rather indicate the limitations and weaknesses inherent in human memory systems. The "sins" are generally categorized into three types: those of omission (forgetting), commission (distortion), and persistence. Here’s a detailed look at each:

1. **Transience**: The tendency for memory to degrade over time. This is the most common form of memory forgetfulness, often manifesting as fading details of a specific event as time passes. It’s a natural process of the brain's adaptive filtering system to prioritize relevant information.
2. **Absent-mindedness**: This sin involves forgetting caused by lapses in attention. When we fail to focus properly at the time of encoding information, the memory does not form strongly, leading to forgetful moments such as misplacing keys or forgetting appointments.
3. **Blocking**: Often experienced as the frustrating "tip-of-the-tongue" phenomenon, blocking is the temporary inability to retrieve a memory, which, while still stored in the brain, cannot be accessed for a short period of time.
4. **Misattribution**: This involves assigning a memory to the wrong source, leading to confusion about where a certain piece of information was learned or the specifics of an event. Misattribution can lead to false memories when one incorrectly combines elements of several events.
5. **Suggestibility**: This sin involves the vulnerability of memory to the power of suggestion, where memories can be altered based on new information or leading questions. This is a significant issue in eyewitness testimonies where suggestibility can lead to inaccurate accounts of events.
6. **Bias**: Memory biases are distortions that pull our recollections in line with our current beliefs, feelings, and what we know about the world. For example, someone might remember past events in a way that reinforces their current attitudes or stereotypes.
7. **Persistence**: This refers to the unwanted recollections that people can’t forget, often involving traumatic or disturbing events. This sin can lead to significant distress, as seen in disorders such as PTSD where the person re-experiences the traumatic event repeatedly.

Together, these "sins" illustrate not just the flaws in our memory system but also its complexity and vulnerability. They highlight the fact that while memory is an incredibly powerful and largely reliable tool, it is not infallible and can sometimes lead to errors that have profound implications for understanding reality and our personal beliefs.

**Fundamentals of Memory Formation**

The process of memory formation and recall involves various mechanisms, each of which plays a critical role in how memories are encoded, stored, and retrieved. Here's an explanation of the key aspects you mentioned:

1. **Working Memory and Association:**
   * Explanation: Working memory acts as a temporary holding space where new sensory information is processed and encoded. It plays a crucial role in forming new memories by associating new information with existing knowledge. This process involves actively manipulating and organizing incoming data by linking it to what we already know, which helps in deeper encoding and easier retrieval later.
   * Significance: The ability to associate new information with existing knowledge allows for more complex thought processes and learning. It enables us to understand new concepts by relating them to previously acquired frameworks, enhancing both recall and comprehension.
2. **Emotional Arousal and Encoding:**
   * Explanation: Emotional arousal during an event can significantly enhance the strength and vividness of the resulting memory. This happens because emotional experiences often involve the amygdala, an area of the brain that facilitates encoding memories at a deeper level.
   * Significance: Emotionally charged memories are usually easier to recall and are more detailed. This mechanism likely evolved because remembering emotionally significant events (like threats or rewards) could be crucial for survival.
3. **Attention and Sensory Memory:**
   * Explanation: Sensory memory acts as the first filter for the countless bits of information we encounter, but only the details that capture our attention are passed on to short-term (working) memory. Attention determines what information makes it past this initial filter and gets encoded into longer-term memory.
   * Significance: By filtering information based on what we pay attention to, our brain conserves resources, focusing on what is most relevant or stimulating. This selectivity is crucial for efficient cognitive functioning, preventing overload from unnecessary data.
4. **Repetition and Rehearsal:**
   * Explanation: Repetition involves the continuous activation of the same neural pathways associated with a particular memory. Rehearsal is a related process where information is consciously reviewed to reinforce the memory trace.
   * Significance: Both repetition and rehearsal are critical for consolidating memories, transitioning them from short-term to long-term storage. This is particularly important in academic settings, where sustained repetition can significantly improve recall of complex information.
5. **Semantic Encoding:**
   * Explanation: Semantic encoding is about processing and encoding information based on its meaning rather than just the surface characteristics. By focusing on the meaning of words, concepts, or events, the information is more likely to be integrated into the existing web of knowledge, making it stronger and more accessible.
   * Significance: Semantic encoding enhances long-term memory because it ties new information to already established cognitive schemas. This deep processing leads to more robust and durable memory formation, facilitating easier access and retrieval of information when needed.

Together, these processes illustrate how different factors and mechanisms contribute to the robust encoding and storage of memories, highlighting the complex interplay between cognitive functions that enables us to learn, remember, and adapt based on experiences.

**Our Memory Model for Reproduction**

1. Attention and Sensory Memory: Connecting Audio-Visual Data for Reproduction
   * How It Works: During the initial phase of memory processing, sensory memories are created almost instantaneously as we encounter audio and visual stimuli. The selective attention mechanism directs cognitive resources towards stimuli that stand out due to their intensity, novelty, or personal relevance, allowing them to be transferred into working memory for more elaborate processing.
   * Connection to Memory Reproduction: For reproduction, we take audio-visual data which is then enhanced by multisensory integration.
2. Working Memory and Association: Leveraging Personal Data for Robust Reproduction
   * How It Works: In the working memory, the focus is on processing and manipulating the information that has passed through the sensory memory. During this stage, new information is integrated with existing knowledge, creating associations that anchor the new memories to previously established cognitive schemas.
   * Connection to Memory Reproduction:  For reproduction, we train on deep personal data to recreate characters, cotext etc.
3. Emotional Arousal and Encoding: Enhancing Vividness through Emotional Charge
   * How It Works: Emotional experiences trigger the release of neurotransmitters that promote strong encoding of memories. The brain prioritizes emotional content, which is processed deeply, making such memories more vivid and durable. Sentiment analysis can be applied to evaluate the emotional content of the memories being encoded, highlighting those with significant emotional impact.
   * Connection to Memory Reproduction: We do sentiment  analysis of the memory and assign value
4. Semantic Encoding: Strengthening Memory Through Meaning
   * How It Works: This encoding strategy involves focusing on the meaning behind the information rather than superficial details. Semantic encoding transforms raw sensory data into meaningful content by linking it to concepts already known, thus embedding it within the network of long-term memory.
   * Connection to Memory Reproduction: Our process is iterative and, therefore, allows an ability to come close to the real memory
5. Repetition and Rehearsal: Solidifying Memories
   * How It Works: Repetition involves the continual reactivation of neural pathways associated with specific memories, while rehearsal is the conscious practice of recalling these memories. Both processes help to strengthen the neural connections that encode the memory, moving it from short-term to long-term storage.
   * Connection to Memory Reproduction: The more repetition in the story the more repetition in the image

Together, these processes create a dynamic system where attention, emotion, personal relevance, semantic depth, and rehearsal work in concert to encode, store, and retrieve memories effectively, allowing for accurate and vivid memory reproduction.

A diagram of a memory reproduction

Description automatically generated with medium confidence

| Category | Research Paper | What can it do? |
| --- | --- | --- |
| Text to 3D | [Magic3d: High-resolution](https://research.nvidia.com/labs/dir/magic3d/)  [text-to-3d content creation](https://research.nvidia.com/labs/dir/magic3d/) | Text to 3D generation |
| [Dreamfusion](https://dreamfusion3d.github.io/) | Text to 3D synthesis |
| single  Image to 3D | [One-2-3-45](https://one-2-3-45.github.io/) | Single image 3D reconstruction |
| [wonder 3D](https://github.com/xxlong0/Wonder3D) | Single Image to 3D using Cross-Domain Diffusion |
| texture | [Texture: Text-guided texturing of 3d shape](https://texturepaper.github.io/TEXTurePaper/) | text-guided generation, editing, and transfer of textures for 3D shapes |
| Motion Transfer | [Few-Shot Human Motion Transfer](https://github.com/HuangZhiChao95/FewShotMotionTransfer) | few-shot human motion transfer that achieves realistic human image generation with only a small number of appearance inputs |
| 2D synthesis | [High-Resolution Image Synthesis with Latent Diffusion Models](https://github.com/CompVis/latent-diffusion) | Latent diffusion models (LDMs) achieve new state-of-the-art scores for image inpainting and class-conditional image synthesis and highly competitive performance on various tasks, including text-to-image synthesis, unconditional image generation and super-resolution |
| Photorealistic 3D Editing | [DreamEditor](https://www.sysu-hcp.net/projects/cv/111.html) | Text-Driven 3D Scene Editing with Neural Fields |
| [InseRF](https://mohamad-shahbazi.github.io/inserf/) | generative object insertion in the NeRF reconstructions of 3D scenes |
| Human Motion Generation | [MDM: Human Motion Diffusion Model](https://guytevet.github.io/mdm-page/) | Natural and expressive human motion generation |
| Misc | [Generative Gaussian Splatting](https://dreamgaussian.github.io/) | 3D content generation framework that achieves both efficiency and quality simultaneously |
| [SEEAvatar: Photorealistic](https://yoxu515.github.io/SEEAvatar/) | text-to-3D avatar generation |
| [Tune-A-Video](https://tuneavideo.github.io/) | Text to video generation |
| [Pix2Video: Video Editing using Image Diffusion](https://duyguceylan.github.io/pix2video.github.io/) | Video style transfer: text-guided video editing |
| [Stable Diffusion XL](https://github.com/Stability-AI/generative-models) | High-quality text to image open source model |
| [Stable Video Diffusion](https://github.com/Stability-AI/generative-models) | Extrapolate image to video |

**Literature: Elements of Filming**

* **Point of view  -** placement of the camera, tilt up, play recorder is on the floor, over the shoulder.
* **Image Content**:
  + Characters: Actors or animated figures portraying roles within the scene. -
  + Dad is looking down, sentiment feeling weight-based placement.
  + **face/eye Micro expression**
  + Objects: Props, vehicles, or other items relevant to the setting or action.
  + Settings: Backgrounds, landscapes, interiors, or exteriors where the scene takes place.
  + Visual Effects: Digitally created elements or enhancements to augment the scene.
* **Composition**:
  + Framing: How elements are positioned within the frame, including shot types like close-ups, medium shots, or wide shots.
  + - Depends on sentiment + Focus
  + Focus is based on facial recognition
* **Lighting**:
  + The angle of light - natural
* **Color**:
* Warm
  + Color Palette: The range of colors present within the frame, including primary, secondary, and tertiary hues.
  + Color Temperature: The warmth or coolness of colors within the frame, influenced by lighting conditions and color grading choices.
  + Color Contrast: Differences in color intensity, saturation, or value used to create visual interest and focal points within the frame.
* **Depth of Field**:
* Focus of the story - What are you paying attention to - repetition in the story
  + Focus: The degree of sharpness or blurriness applied to objects or areas within the frame, controlled by camera settings such as aperture, focal length, and focus distance.
  + Shallow Depth of Field: A narrow range of focus, isolating subjects from the background and foreground.
  + Deep Depth of Field: A wide range of focus, keeping multiple elements within the frame sharp and in focus.
* **Movement**:
* Depends on the emotion of the character
  + Camera Movement: The motion of the camera within the frame, including pans, tilts, zooms, dollies, or tracking shots.
  + Object Movement: The movement of physical objects or elements within the frame, such as characters walking, vehicles moving, or environmental changes.
  + Character Animation: The actions and gestures performed by characters within the scene convey emotion, personality, and narrative progression.
* **Sound**:
* External sound mentioned
  + Dialogue: Spoken words exchanged between characters, conveying information, emotion, and character relationships.
  + Sound Effects: Audio elements representing actions, events, or environmental sounds within the scene, enhancing realism and immersion.
  + Music: Score or soundtrack accompanying the scene, setting the mood, pacing, and emotional tone.
  + Ambient Noise: Background sounds or atmospheres contributing to the sense of place and atmosphere within the scene.

Stakeholders (including who loses out):

People who feel nostalgia, who write diaries and who like verbally sharing personal stories and at the same time open to new technologies would be the main users of interest.

There might be groups of people who are against using generative technologies to blur the line of real and synthetic imaging, similar to the groups of people who are against OpenAI products. There might be a mixed perception among the film production industry as well, some would go against using generative AI for virtual production.

IP

The copyright of memory and stories people share with us will solely belong to individuals. And individuals have the rights to license the contents to others for reproduction.

“Act One” Implementation

The team now has a good mix of expertise with film making and engineering. The space of multimodal generation is drawing overwhelming research attention so it is an exciting domain to work on for many talents. The major hurdle though is two-fold:

1. Training models require large capital funding and stocks of GPU resources. As a starter, using on-demand cloud services would require less capital upfront till MVP is reached.
2. It can be hard to recruit the top-talents in this research field. So collaborating with those who have a strong interest would be easier.

Operating model:

If we can reach an MVP that resonates with target user segments, there is a chance of world-of-mouth marketing. Though reaching break-even would be challenging as many startups in this field have not been able to do so.

Acts Two, Three and beyond:

We are seeing a shift of product focus in generative modeling, from generative platform to full-scale storytelling toolkit. But it is too early to tell what the operation and sustainable business model will look like.