# Human Augmentation Platform using Context, Biosignals and Language Models

## Abstract

There is disclosed systems and methods for human agency enablement through an integrated use of context information, historical work product, biosensors, explicit user input and a large language model. The system comprises input means, tokenization, a large language model and an output stage capable of enabling agency using output tokens from the language model.

## System Overview

Diagram

Description automatically generated

Diagram

Description automatically generated

Figure 1: Overview of user agency augmentation system.

In an exemplary embodiment, a user interacts with a mobile or wearable computing device as shown in the figure above. This embodiment provides the user with capability augmentation by utilizing biosignals, inference of the user’s environment and physical state, a range of on-device and off-device sensors, user history, and direct input to prompt a large language model to generate multimodal output(s). These outputs are used to extend user agency into real and virtual endpoints. They include text which can be converted to speech by a text-to-speech model, text presented on the user’s or their companions’ mobile or wearable computing device, and a multimodal generative AI output like a picture and sound.

Figure 1 shows the pathways signals take from input, by a sensing device or the user, to output in the form of text, text-to-speech, or multimodal artifact. Each step in the process and possible embodiments are described below.

### Context subsystem

Diagram

Description automatically generated

Figure 2 …

The context subsystem is responsible for generating a context prompt that represents the **Context** of the user to be passed to the final prompt composer for integration. Broadly speaking ***context*** consists of sets of tokens generated from a variety of different **data sources** including; sensor data captured on or around the user, background material that provides information about the user’s previous history, and information about the current task or interaction the user may be engaged in. Each *data source* generates a set of preliminary tokens which are then passed to all other data sources to be consumed as input for the final tokenizers for each data source. Each data source refines its output based on the ***context*** provided by other data sources, which is particularly important for background material (below). For example, the ***context*** used by the Final Background Material tokenizer to determine which background material elements are likely to be relevant is the set of tokens generated by the all raw data source tokenizers (which may (for example) have identified an individual person in the environment).

**Application context**

In some instances, the user is interacting with an application on a computing device. In these instances explicit specification of the application can greatly enhance the context subsystem’s knowledge of the user context and facilitate a more optimal context token set.

**Background material** may be plain text, a structured database, cloud data storage (structured or unstructured) or any mixture of these data types. In one embodiment *Background Material* may include textual descriptions of activities that the user has done in a similar context and their prior outcomes, if relevant. In other embodiments, the a structured or unstructured database of *Background Material* will include general information about the user, about topics relevant to the user’s current environment, conversational history (ies), notes or any other material related to the user’s situation which is of a contextual or historical nature. Background material In some embodiments, the background material is first converted into a plain text stream and then tokenized using a plaintext tokenizer.

For structured historical data such as plaintext, database or web-based textual content, tokens consist of the numerical indexes in a vectorized (e.g. word2vec or similar) representation of the text content. In some embodiments, an unsupervised neural network technique called an autoencoder could be utilized to transform plaintext inputs into high dimensional vectors that are suitable for indexing and tokenization ingestion by the prompt composer.

For example (<https://platform.openai.com/tokenizer>):

Graphical user interface, text, application

Description automatically generated

Below is the list of numerical indices corresponding to these words:

A picture containing application

Description automatically generated

The context subsystem may also include a mixture of physical sensors such as microphones and cameras that connect with network-connected and embedded data sources and models to generate a numerical representation of a real-time context estimate.

An example of a background/context prompt is shown below:

For the rest of this conversation, you are a language model that is communicating to a user diagnosed with ALS. The person with ALS is sending messages to you, and would like you to take what they say to you and expand it into a full utterance. They use prediction to compose the language they are sending you whenever it is available and accurate, but uses an eye gaze or BCI system that only allows one interface interaction per second. In that system they will always make choices that are intended to produce grammatically correct sentences. When they are finished writing their message, it is sent to you. Respond to each message with the three most likely full utterances that the user intended. Do not include explanations; the string you respond with will be spoken by their speech generating device. This is the user:

* Sarah, a 60-year-old retired school teacher, is sitting in her garden, basking in the warm sunlight and enjoying the peaceful sounds of birds chirping. She has just finished reading a mystery novel recommended by a close family member or friend, feeling content and relaxed. As a literature enthusiast, Sarah has a long history of discussing books with her loved ones.
* Relevant Context Data:
  + Conversation history: Recent discussions about books with family and friends, including favorite authors, genres, and specific titles
  + Personal preferences: Fondness for mystery novels, historical fiction, and biographies; appreciation for strong character development and engaging plots
  + Language corpus and demographics: 60-year-old, retired school teacher, well-versed in literary terms and expressions
  + Mood: Content, relaxed, and eager to share her thoughts on the novel
  + Relationship: Close family member or friend with shared interests in literature
  + Environmental audio: Sounds of birds chirping, rustling leaves, and distant neighborly conversations in the garden
  + Front-facing camera: Images of blooming flowers, lush greenery, and the mystery novel's cover
  + Network state, motion, and positioning: At home, sitting in a comfortable garden chair
  + Reading preferences: Mystery novels, historical fiction, biographies, and classic literature
  + Educational background: Years of teaching experience in literature, familiarity with various literary periods and styles.

The user does not use punctuation, but add it automatically where predicted. The punctuation if added should reflect the mood and tone where relevant. Insert pronouns, prepositions, Wh- words, and and other high-frequency terms automatically in his output. If the user sends an emoji, use it to topically or thematically improve prediction and alter tone. Use all contextual information to weight your responses, and weight prior conversation history. After each input review the prior inputs and weight your subsequent predictions based on the context of the thread. Remember, for the rest of this conversation you are a large language model assistant who replies only with the three top predictions of what the user intended to say with the message they sent. Do not add any other information. If you understand, please respond with “proceed.”

In some embodiments, context also includes audio, visual or other multimodal data. In this case tokenization may be performed using a convolution-based tokenizer such as a vision transformer (e.g. <https://www.researchgate.net/publication/359647022_Multimodal_Fusion_Transformer_for_Remote_Sensing_Image_Classification>). In some alternate embodiments, the multimodal data is quantized and converted into a token using a codebook (see <https://vaclavkosar.com/ml/Tokenization-in-Machine-Learning-Explained>). In yet other alternate embodiments, multimodal data is directly encoded and presented to the language model as a vector space encoding. An exemplary system that utilizes this tokenizer strategy is Gato (<https://arxiv.org/pdf/2205.06175.pdf>), which can ingest a mixture of discrete and continuous inputs, images and text as tokens (see section 2.1 of the paper).

The final output from the context subsystem is a sequence of text tokens containing the combination of the background, audio/video and other context token sequences.

## Biosignal Subsystem

Diagram

Description automatically generated

The biosignal subsystem consists of a mixture of physical sensors on or near their body that connect with network-connected and embedded data sources and models to generate a numerical representation of a biosignal estimate. The biosignal tokenizer encodes the biosignal estimate with associated data to generate at least one biosignal token sequence.

In some embodiments, the mobile or wearable computing device has a set of sensory peripherals designed to capture user biometrics. In some embodiments, these biometric sensors include some combination of EEG, ECoG, EMG, ECG, pulse, heart rate variability, blood sugar sensing, etc. These biometric data are converted into a biosignal token sequence in the biosignals tokenizer step.

Biosignals may be tokenized through the use of a classifier. It is common practice for a biosignal raw signal data to be analyzed in real time using a classification system. For EEG signals, a common choice is CCA (canonical correlation analysis) which ingests multi-channel time series EEG data and outputs a sequence of classifications corresponding to stimuli that the user may be exposed to. In one example, these signals may consist of steady state visually evoked potentials (SSVEP) which occur in response to specific visual stimuli. In other possible embodiments, the classification may consist of a binary true/false sequence corresponding to a P300 or other similar neural characteristic. In yet other possible embodiments, the classes may consist of discrete error related potential (ERP) responses. It will be clear to the reader that other biosignals including EOG, EMG, EKG may be similarly classified and converted into symbol sequences. In other embodiments, the signal data may be directly tokenized using discretization and a codebook.

The biosignals subsystem may also have a connection with the context subsystem in advance of any prompt composition. This connection may bidirectionally inform each of the subsystems to enable more precise, or more optimal token generation.

### User Input Prompt

In one embodiment, the user provides an additional token sequence in one or more sensory modes which can include a sequence of typed or spoken words, an image or sequence of images, and a sound or sequence of sounds. The biometric and optional multimodal prompt input from the user are tokenized using equivalent techniques as for the context data.

Prompt Composer

A prompt composer consumes the context, biosignals and user prompt tokens and constructs a single prompt token in a format suitable for a large language model (LLM) such as GPT-4 (ref. OpenAI) along with tokens that identify the required output modality. In an alternate embodiment, the prompt composer further generates an embedding which can be provided separately to the LLM and used in an intermediate layer of the LLM.

In an exemplary use case, the context subsystem may generate a token sequence corresponding to the plaintext: *‘The user has travelled to Los Angeles to visit a doctor specializing in rare diseases. The user is sitting in the doctor’s office and preparing to discuss their disease. The user is looking at the doctor who is has just asked the user for an update on their condition.*’. The Biosignals prompt may include a token sequence corresponding to the user selecting ‘speak’ with a computing device to select this directive using an EEG based brain computer interface. The User Input Prompt may include a token sequence corresponding to the plaintext ‘The user has selected ‘summarize my recent disease experience’. In the case, the prompt composer simply appends these three token sequences into a single token sequence and passes it to the large language model. In an alternate embodiment, the prompt composer replaces the Biosignals prompt with a token sequence corresponding to the plaintext ‘Generate output in a format suitable for speech synthesis.’

The LLM consists of a pre-trained machine learning model, such as GPT, which takes as input the token sequence generated by the prompt composer and generates a token sequence output that can be converted back into plaintext, or which can be consumed by a user agency process directly as a token sequence. In an alternate embodiment the output of the LLM further constitutes embeddings which are decoded into multimodal or time-series signals that can be utilized by agency endpoints. Once determined, the output is digitally communicated to an agency endpoint.

In an alternate configuration, the system contains multiple LLMs, each of which is pretrained on specific application, context or agency domains. In this configuration, the context subsystem is responsible for selecting the appropriate LLM(s) for the current estimated user context. In some embodiments, [mixture-of-expert](https://proceedings.mlr.press/v162/du22c/du22c.pdf)s models like [GLaM](https://proceedings.mlr.press/v162/du22c/du22c.pdf) can be used for this. In some embodiments, models can be fine-tuned by the user. For example, the user could provide an LLM classifier model with exemplars of classes either by speaking or writing them, and through [few-shot learning](https://arxiv.org/abs/2209.01390), the model can improve accuracy.

In one embodiment, the selected user agency process is a speech synthesis system capable of synthesizing token sequence as spoken language in the form of a digital audio signal. In another embodiment, the endpoint is a text composition interface associated with a communication application such as email, social media, chat, etc. In a further embodiment, the output is a multimodal artifact such as a video with text, a robot command sequence, etc. In another embodiment, the output can be dynamic alteration of a user’s interface, access method, or complexity of interaction to maximize utility in context. In another, the system’s output could be constrained to a subset of domain-relevant utterances for applications such as employment, industry, or medical care.

In some embodiments, the output is additionally encoded using an encoder framework such as an autoencoder. In this system, the output of the encoder framework may be a sequence of control commands to control a robotic system such as a powered wheelchair, prosthetic or powered exoskeleton. In one version of this embodiment, the prompt includes either biosignals or user input tokens which represent the user’s desired configuration and the language output includes detailed steps that a robotic controller can digest, once encoded. In this embodiment, the user may express a desire to move from location A to location B and the combination of the LLM and the robot controller will generate an optimal path as well as a detailed control commands for individual actuators. In a further embodiment, the LLM may generate a novel control program which is encoded by parsing or compiling for the target robot control platform.

In general, the system can be viewed as a kind of application framework that uses the Context/BioSignals and User Input Prompt sequences to facilitate interaction with an application much like a user would use their finger to interact with a mobile phone application running on a mobile phone operating system. Unlike a touchscreen or mouse/keyboard interface, this system incorporates real time user inputs along with an articulated description of their physical context and historical context to facilitate extremely efficient interactions to enable user agency.

## Other exemplary Embodiments

A system and method for improving indoor navigation is provided. The system includes a biosensor subsystem that can sense physical signals, such as heart rate and respiratory rate, and a context subsystem that infers the user's intent by utilizing contextual information such as lidar positioning and indoor map data. The system further includes a pre-trained natural language model that can utilize the set of tokens to provide personalized indoor navigation instructions and alerts to the user. In this embodiment, the context subsystem may utilize a set of cameras mounted to a user’s wheelchair to construct a token sequence corresponding to their visually available surroundings. The Background material may consist of floorplans or other information about routes or preferred access methods for the user’s indoor location. The biosignals subsystem may provide a go/no-go token to the system depending on the user’s estimated anxiety and/or their explicit attention to a visual stimuli. The user input prompt may include a specific user directive such as ‘drive to nearest bathroom’.

A system and method for controlling IoT devices with eye blinks is provided. The system includes a biosensor subsystem that can sense eye movements, and a context subsystem that infers the user's intent by utilizing contextual information such as blink patterns and device locations. The system further includes a pre-trained natural language model that can utilize the set of tokens to provide personalized device control and automation to the user. In this embodiment, the context subsystem may utilize Bluetooth connectivity to sense local IoT devices. The background material may consist of previous commands or protocols. The biosignals subsystem may provide tokenized eye-blink data. The user input prompt may include specific IoT commands or partial commands.

A system and method for identifying conversation participants in a crowded environment is provided. The system includes a Bluetooth antenna that can detect other devices in close proximity, and a context subsystem that infers the user's intent by utilizing contextual information such as speech patterns and device identities. The system further includes a pre-trained natural language model that can utilize the set of tokens to provide personalized conversation management and social interaction support to the user. In this embodiment, the context subsystem may utilize Bluetooth connectivity to sense local IoT devices. The background material may consist of previous conversations or detail about conversation participants. The biosignals subsystem may provide tokenized fatigue/emotion data. The user input prompt may include keywords or phrases related to desired speech output.A system and method for incorporating personal multimedia artifacts into the LLM corpus is provided. The system includes a media ingestion module that can import personal multimedia artifacts, such as photos and videos, into the LLM corpus, and a context subsystem that infers the user's intent by utilizing contextual information such as artifact metadata and user preferences, which constitute background data. The system further includes a pre-trained natural language model that can utilize the set of tokens to provide personalized multimedia content generation and sharing to the user. In this embodiment, the biosignals subsystem may provide tokenized spatial attention and emotion data. The user input prompt may include names of files and objects in user’s multimedia artifact collection, as well as commands for fine-tuning models.

An AR communication app that predicts or estimates emotions based on facial expressions, body language, and biosignals, enhancing communication for individuals with social communication difficulties is provided. In this embodiment, the context subsystem utilizes phone sensor data, biosignals, and machine learning algorithms to detect the user's emotions and provide real-time feedback on their conversation partner's emotional state. The app is valuable in contexts where individuals with social communication difficulties require additional support in understanding their conversation partner's feelings. The background material may include data about previous interactions with conversation partners. The system can provide personalized feedback to the user, helping them to adjust their communication style to better connect with others. In one embodiment, the output of the language model is a set of suggestions that are rendered into private audio for the user. In other embodiment, the output is visually rendered into a heads-up display. In yet another embodiment, the output is rendered into a multimodal feedback that expresses an affect grid coordinate to the user using a mixture of output modalities (e.g. valence is mapped to an auditory pattern and arousal is mapped to a haptic pattern).

A system and method is provided for enhancing reading comprehension by detecting the user’s cognitive load through biosignals and offering real-time explanations, definitions, or alternative phrasing for difficult content. In this embodiment, the system includes a biosensor subsystem capable of sensing at least one physical and/or mental signal, a context subsystem that infers the user's cognitive load utilizing as the biosignals, time, location, and other contextual information processed into tokens, and a pre-trained natural language model that can utilize the set of tokens to provide real-time explanations, definitions, or alternative phrasing for difficult content. The feedback can be communicated to the user via an audio means or a mixed reality visual display that overlays explanations or hyperlinks on the reading materials. In another embodiment, the feedback can communicate the user’s mental state back to the user via a multimodal sensation in order to enable conscious recognition of their mental state.

A language processing system for autonomous vehicles that utilizes biosignals to enhance safety and comfort for passengers. The system includes sensors that capture biosignals such as heart rate and respiratory rate, as well as a language model that processes contextual data from the vehicle's surroundings and prompts from the passengers. The system provides real-time feedback to the vehicle's autonomous control system to adjust speed, route, and other factors to improve passenger safety and comfort. The output of the language model is converted into multimodal sensations through vehicle instruments such as the steering wheel, driver heads up display and/or over the in-vehicle audio system.

A language processing system for virtual e-commerce platforms that utilizes biosignals to personalize shopping experiences for users. The system includes sensors that capture biosignals such as heart rate variability and skin conductance, as well as a language model that processes contextual data from the user's browsing history and prompts from the user. The system provides real-time feedback and guidance to the user to improve their shopping experience, including personalized product recommendations and checkout optimization. The personalized feedback generated by the LLM is converted into visual, auditory or haptic means to be rendered to the user.

An augmented reality (AR) system for travelers that utilizes biosignals and language processing to enhance their travel experience is provided. The system includes a wearable device that captures biosignals such as heart rate and respiration rate, as well as a language model that processes contextual data from the user's surroundings and prompts from the user. The system automatically translates language between the user and their conversation partner(s) and also adapts it to their knowledge and state. The system provides real-time feedback and guidance to the user to improve their travel experience, including personalized recommendations for activities, dining, and accommodations. Feedback can be provided using visual, auditory or haptic transformations of the LLM output.

A system and method is provided for providing personalized rest and activity recommendations to optimize energy levels throughout the day by predicting the user's fatigue levels based on biosignals, phone sensor data, and past behavior. In this embodiment, the context subsystem , utilizes phone sensor data such as movement and usage patterns and the biosignals subsystem can sense at least one physical and/or mental signal, including but not limited to biosignals such as heart rate, skin conductance, and muscle tension. The biosignals subsystem in this embodiment may also infer the user's fatigue level utilizing EEG, heart rate, and similar biosignals. A pre-trained machine learning model may then utilize the set of tokens to generate personalized rest and activity recommendations for the user to optimize their energy levels throughout the day or progress over the long term. The system is valuable in contexts where individuals require support to manage their fatigue levels, such as individuals with chronic fatigue syndrome or those in high-stress work environments, or distance athletes. Feedback can be provided using visual, auditory or haptic transformations of the LLM output.

A system and method is provided for optimizing XR workspace productivity by predicting the user's cognitive load with tokenized biosignals and adjusting the workspace layout, interface elements, and notifications accordingly. In this embodiment, the system includes a biosensor subsystem capable of sensing at least one physical and/or mental signal, including but not limited to biosignals such as heart rate, skin conductance, and muscle tension.

The system also includes a context subsystem that infers the user's cognitive load utilizing contextual information such as biosignals and other input processed into tokens. The LLM can be prompted to provide computer interpretable codes that modify or change the layout of the workspace. The system is valuable in contexts where individuals require support to manage their cognitive load in virtual workspaces, such as remote workers or individuals with cognitive disabilities.

A language processing system and method for social media platforms that utilizes biosignals to enhance user engagement. The system includes sensors that capture biosignals such as skin conductance and facial expressions, as well as a language model that processes contextual data from the user's social media history and prompts from the user. The system monitors social media activity and provides real-time feedback and guidance to the user to improve their engagement with social media, including personalized recommendations for content and real-time feedback on their emotional state as well as suggestions for alternate phrasing of posts. This feedback can be provided to the user as a visual overlay. In some embodiments the feedback is provided via transformation of the of biosignals and a valence estimate based on the user’s generated content. This feedback may be in the form of a visual, auditory or haptic sensation.

A system and method for enhancing safety in the workplace by monitoring biosignals and phone sensor data to detect signs of fatigue or distraction in employees.

The context subsystem utilizes wearable or handheld sensor data such as movement and usage patterns along with the biosignals subsystem to sense at least one physical and/or mental signal, including but not limited to biosignals such as heart rate, skin conductance, and muscle tension to detect fatigue or distraction. The system uses a large language model to provide alerts, prompts, or recommendations for breaks, rest, or changes in work conditions in augmented reality to reduce the risk of accidents or errors. This feedback is provide in the form of visual, auditory or haptic stimulation.

A system and method for reducing social isolation and loneliness by monitoring biosignals and other sensor data to detect and predict changes in the user's mood, activity, and social behavior. A historical record of the user’s biosignals and context is tracked and fed into a l= a large language model to provide personalized recommendations, activities, or connections to support the user's social well-being. This feedback is provided in the form of visual, auditory or haptic stimulation.

A system for enhancing the user's experience in a museum, or other public of novel location using a combination of AR and a language model. The system utilizes sensor data, user language history, search history and interests to drive the prompt composer. The system then generates real-time descriptions, translations, or explanations of the exhibits, guiding the user to areas of particular interest and relating exhibits to their existing knowledge. This feedback is filtered and prioritized based on user location, gaze direction as well as estimated interest and arousal. This feedback is provided in the form of visual, auditory or haptic stimulation.

A system for improving the user's experience in a hospital. The system uses AR and a language model to label medical equipment and provide real-time information, guidance, and instructions. The system also adapts the visual experience to the user's cognitive load and preferences, reducing stress and promoting understanding during medical procedures. This feedback is filtered and prioritized based on user location, gaze direction as well as estimated interest and arousal. This feedback is provided in the form of visual, auditory or haptic stimulation.

A system and method for improving language translation is provided. The system includes a biosensor subsystem that can sense physical signals, such as stress levels and respiration, and a context subsystem that infers the user's intent by utilizing contextual information such as language proficiency and cultural background. The system further includes a pre-trained natural language model that can utilize the set of tokens to provide personalized language translation and interpretation services to the user. These translations can be provided to the user via visual stimuli of an XR display and/or by auditory stimuli.

A system and method for improving driving safety is provided. The system includes a biosensor subsystem that can sense physical signals, such as heart rate and respiratory rate, and a context subsystem that infers the user's intent by utilizing contextual information such as driving conditions and distractions. The system further includes a pre-trained natural language model that can utilize the set of tokens to provide personalized driving safety recommendations and alerts to the user. These translations can be provided to the user via visual stimuli of an XR display, by auditory stimuli or by haptic feedback through vehicle instrumentation.

### Dialog system that only uses biosensing to drive responses

In another embodiment, the system utilizes context – including from biosignals and other sensor input - to provide input to the prompt composer. This embodiment is useful for driving agency outputs that are indicative of user body language for example. In one use case, a user is engaged in dialog with another person. As the other person talks, the audio is converted into tokens using the context tokenizer and the user’s physiological and cognitive response is estimated from biosensor data. These two tokenized streams drive the language model to generate agency that expresses the user’s likely responses to the speaker. The user can then select appropriate responses and have them communicated to the dialog counterparty. In this embodiment, no user prompt is required to drive predictions. In another embodiment, ‘nodding head and listening with interest’ could be an output from the language model, which could then be used to drive some robotic system or simply an emoji response that is rendered back to the speaker. In another embodiment, a user’s physical state could be used to prioritize utterances that would address their immediate needs or activities, such as requesting assistance or exposing health monitor information to the user when exercise activities are detected.

### Biosignals are used to infer the tone or style of the model output

In another embodiment, the biosignals tokenizer utilizes biosignals information, in part, to generate stylistic or tonal prompt tokens. In one exemplary embodiment, the system can use brain sensing data and/or heart rate data to estimate the user’s level of arousal or engagement with the context. When arousal is high, the biosignals tokenizer can generate more emphatic or excited tokens: e.g. ‘excited, happy, engaged’. When arousal is low, the tokenizer can generate tokens with less interest or excitement: e.g. ‘bored, disinterested.’ The system could also be directly adjusted by the user, such as selecting emoji to alter the output’s persona, tone, or “mood.” These prompt tokens are then incorporated into the prompt via the prompt composer and generate outputs from the LLM that reflect the user’s sensed tone and style. These outputs could be rendered visually or auditorily or in some embodiments, could be transmitted over a network to a remote server.

In another embodiment, the context system can sense the state of a controllable entity such as a vehicle, assembly line or robotic system. A tokenized representation of this state can be used as input to the LLM along with a prompt identifying the desired state of the controllable entity. The LLM can be instructed to generate corrective control commands that will modify the state of the controllable entity. In one exemplary embodiment, the context subsystem may serve assembly line robots by detecting anomalies in the visual field that indicate changes in orientation of items in the assembly line.

Simultaneous Users

In some embodiments more than one user is interacting with a common artifact, environment or in a social scenario. In these embodiments, each user may have an instance of one or more of the embodiments described herein. Further when multiple such systems are present, they may establish direct, digital communication with each other via a local area or mesh network to enable direct context transmission and exchange of LLM outputs. In some instances one or more of the simultaneous users may be a robot or other autonomous agent. In yet other instances one or more of the users may be an assistive animal such as a sight impaired support dog.

### Background is the user’s collected works in a specific topic area

In another embodiment, the historical work product is a collection of documents, images or other media that represent the user’s collected works in a specific topic area. For example, the user might be an expert in a specific topic such as eighteenth-century literature, or electrical engineering. The user may have written articles, curricula, videos, etc. These can be incorporated into the historical work product and summarized or tokenized as appropriate. In some embodiments, the amount of background work is substantial and could alternatively require that the LLM itself is retrained with this additional information. This historical information is incorporated into the context token stream and provided as background to the LLM prompt composer. In this configuration, the sensed context, biosignals and user prompt are used to construct a prompt that specifically relates to the historical work product. In one use case, this would cause the output text to be in the style of the background material or framed in the technical jargon of a specific technical discipline.

### Historical user works pretrain the LLM to be better respond to compositional prompts

In another embodiment, the historical compositional data is pre-tagged with relevant semantic context tokens prior to any utterances being generated and is passed as training input to the LLM. As in the above embodiment the LLM is tuned so that it is capable of generating output that contains more user specific knowledge and tone appropriate to the target output format. For example, a user might import their chat history which would be tagged not only with the conversation partner but also with the type of conversation, such as an informal one about sports, which could then be used to improve the models ability to generate speech in the users tone when they discuss sports with friends or family in the future.

### Background is a specific set of tasks or instructions for the user to execute a task

In another embodiment, the historical work product is a specific set of instructions related to a task that the user needs to perform or will soon perform. The task can be in nearly any domain including cooking, mechanical repair, craftmaking or computer programming. The context subsystem estimates the appropriate state of the target of the task using available sensors. The biosensor information estimates the user’s body configuration as well as their level of mental workload, fatigue and frustration with the task. The user prompt can include user instructions for pacing the task execution. The prompt composer can create a prompt that asks the LLM to generate instructions for the next step in the process and/or for the system to provide feedback on the user’s performance of the task. In some embodiments, this feedback can also include coaching to improve user confidence or reduce anxiety. The instructions or coaching feedback may be provided to the user as visual or auditory stimuli using an appropriate rendering device such as a pair of XR glasses or a headphone.

### Context estimate is the difference between observed and expected

In another embodiment, the context subsystem uses available sensors and background information to estimate the difference between the observed state of an external artifact and a desired or expected state. For example, the user may be repairing an engine but the context subsystem notes that the user is missing some key tools to complete the task. The context subsystem may drive the prompt composer to ensure that the user has the correct tools for the job at hand, e.g. ‘you’ll need a screwdriver with a T6 bit to complete this repair’. In another use case, the context estimator may use external sensors to identify a potentially hazardous situation such as collapsed roadway or pathway that a user is walking along. This context discrepancy can be used to provide feedback to the user which is then verified using the biosensing system (e.g. the user stops or changes course).

### Background includes labels that can be associated with sensor data

In another embodiment, the background data includes informational labels that can be associated with sensor streams such as video or audio. These informational labels can be used along with biosensor information such as the look vector of the user’s eyes in order to provide contextually relevant commentary on an environment. For example, the user might be in a museum and the context system’s video sensor classifies a specific artifact for which there is background information. This can then generate a running monologue as a user explores the museum environment. Another embodiment is expressive, where a forward-facing camera could detect a person, object, or activity that enriches the model’s potential output. In one exemplary embodiment a vision impaired user may receive a running monologue of key classified elements of their environment in order to enable them to navigate or otherwise function in an unfamiliar location.

### Biosignal tokenizer includes information about desired vs observed user body state

In another embodiment, the biosignal subsystem specifically identifies differences between observed physiology or cognitive state and desired or expected state. This information is encoded and combined with the context and multimodal prompt to drive a coaching or guidance oriented output from the LLM. In a use case, a user’s blood sugar may be higher than expected and the user may be in the process of selecting food for lunch (context). The LLM may provide coaching feedback to the user indicating that they should make a healthy selection. In another, a forward-facing camera may detect a restroom sign and inform a user with a disability that it’s available if they have been seated for an extended period of time.

### Biosignals and semantic context detection of turn taking and request for a turn



Figure # Annotated sample embodiment with turn taking detection

In another embodiment the biosignals subsystem, context subsystem, and the LLM will work together to detect whether or not it is time for the user to respond to the speech of a conversation partner. The biosignals subsystem will use brain sensing to determine if the user has detected or is anticipating a question or declination in speech which they are expected to respond to. The microphone will record the conversation partner’s speech for use in determining the appropriate response. At an experimentally determined threshold level of brain sensing anticipation and microphone silence the conversation partner text will be added as relevant background material and the brain sensing data will be sent to the LLM. The LLM will then take this input and generate responses that are tonally, semantically, and modality appropriate to the context provided (including but not limited to the new anticipatory brain sensing data, the speech to text from the conversation partners microphone, and the rest of the conversation history).

### Using Biosignals information to infer adequacy of LLM output

Figure X# A simplified version of the biosignals feedback system for detecting whether or not the output of large language model should or should not be automatically rejected

In another embodiment, after the LLM model has generated a suggested item, biosensors will be used to detect whether a potential error/surprise is detected. Error related potentials (ERPs) are known to signal user surprise when presented with stimuli. Whenever the system detects that error/surprise is detected the user’s action will be recorded in response to the suggestion, whether it is accepted or rejected. The response itself, the strength of the response, and the number of sensors agreeing with the response, will be used in combination with the input tokens to the system (from the original prompt) to feed into an ***unexpected output*** ***machine learning model***. This model will use supervised learning to predict whether or not a given combination of error/surprise response + prompt token can be relied upon to predict whether or not a user will reject or accept a suggestion. If the likelihood of suggestion rejection is sufficient (above a threshold determined experimentally) the system shall automatically reject the suggestions and generate feedback tokens. This automatic rejection feedback is then passed back into the prompt generator to provide negative feedback to the LLM. Examples of biosignal feedback include EEG signals, and facial expressions. In general, in this embodiment one or more biosignals based estimates of user surprise are used to modify or refine an existing machine learning model, including for example, an LLM.

For example the model may generate an utterance that is positive in tone when the user was expecting a message with a negative tone, this incongruity would be detected and a new prompt would be generated that includes a rejection of the statements in the current prompt.

Sensor Availability (our hardware + android typical + eye tracking + 3rd party integrations + any other biosignal data

### Structural claims

In the above embodiments, it should be understood by one skilled in the art that a record of inputs and responses can be used to retrain and enhance the performance of any of the system components. For example, a record of natural language outputs could be scored based on some external measure and this data then used to retrain the model itself.

All embodiments provide for some type of feedback to a user or an entity. It should be clear that this feedback could be in the form of a sensory stimuli such as visual, auditory or haptic feedback. However, it should also be clear that this feedback could be transmitted over a network to a server which may be remote from the user. This remove device may further transmit the output from the system and/or it may transform the output into some other type of feedback which is then communicated back to the user and rendered as visual, auditory or haptic stimuli.

Some or all of the elements of the processing steps of the system may be local or remote to the user. In some embodiments processing may be both local and remote while in others, key steps in the processing will leverage remote compute resources. In some embodiments, these remote resources will be edge compute while in others they may be cloud compute.

In some of the embodiments, the user may explicitly select or direct components of the system. For example, the user may be able to choose between LLM’s that have been trained on different corpus if they prefer to have a specific type of interaction. In one example, the user may select between an LLM trained on clinical background data or an LLM trained on legal background data. These models will provide distinct output tokens that are potentially more appropriate for a specific user intended task or context.

Proposed Claims:

1. A system and method for generating user agency comprising:
   1. A context subsystem that can infer a user’s context,
   2. A biosensor subsystem capable of sensing at least one physical and or mental signal,
   3. A pre-trained natural language model that can utilize a set of tokens providing prior knowledge along with a multimodal prompt,
   4. An output stage that can transform the language model output into at least one form of agency.
2. The system of claim 1 where the biosensor system includes at least one EEG based sensor.
3. The system of claim 1 where at least one form of agency includes neural stimulation to the user with tDCS or similar.
4. The system of claim 1 where the biosensor system includes at least one brain sensor.
5. The system of claim 1 where the output stage consists of a speech synthesis system capable of generating audible sound.
6. The system of claim 1 where the feedback stage includes at least one EEG based sensor.
7. The system of claim 1 where the feedback stage consists of at least one brain sensor.
8. The system of claim 1 where the output stage is used to drive a generative AI system to create a multimodal artifact.
9. The system of claim 1 where the context subsystem includes an environmental sensor such as a camera or microphone array.
10. The system of claim 1 where the set of tokens are derived, in part, from the biosensor, and context subsystems.
11. The system of claim 1 where the set of tokens are derived, in part, from historical utterances, dialog, email, social media and other multimodal communications of a single user.
12. The system of claim 1 where an optional feedback stage automatically reinforces or rejects the language model output.