conversational Interactions (Simon and Laya)

emotional Computing> tiea jenny lina

generation with LLMs –> Dhruv and combine all

Week 1- state of art. rough summary of the emotional computing papers.

Emotional Computing

Paper 1 — Catania & Garzotto (2020): "Emozionalmente – Crowdsourced Emotional Speech Corpus"

Focus / Contribution

- **Created an Italian emotional-speech dataset using a web-based crowdsourcing platform.**
- **Evaluated human and AI emotion recognition performance (using Wav2Vec 2.0).**

Key Insights

- **Crowdsourced emotional data can be accurate and scalable.**
- **Deep-learning models outperform human listeners in emotion detection.**
- **Cultural and linguistic specificity greatly affect emotion recognition accuracy.**
- **Some emotions (fear, disgust) remain difficult to classify.**

Design & Application Findings

- **Enables automated, real-time speech-emotion feedback.**
- **Systems should tolerate ambiguity and multiple interpretations. -** **Given that even humans misclassify emotions, design should not treat the AI's reading as absolute truth. Instead, feedback could express confidence levels ("You sound slightly anxious, confidence 70%") or offer interpretive reflection ("You may be expressing tension—does that feel right to you?").**
- **Crowdsourcing or user-generated emotion samples can personalize training.**

Paper 2 — Wani et al. (2021): "A Comprehensive Review of Speech Emotion Recognition Systems"

Focus / Contribution

- **Reviews the full SER pipeline from data collection to classification.**
- **Compares traditional machine learning with modern deep learning methods.**

Key Insights

- **Deep models (CNN, RNN, LSTM) outperform traditional approaches.**
- **Dimensional emotion models (valence–arousal) better capture nuance.**
- **Multimodal fusion (speech + text + facial) increases robustness.**
- **SER remains limited by labeling bias, noise, and cross-cultural variation.**

Design & Application Findings

- **Use dimensional feedback rather than fixed emotion labels. -** **Use dimensional feedback rather than fixed labels. Instead of telling users "You sound angry," the system could show a** position on a valence–arousal map (“high arousal, negative tone”), helping users explore the gradient of their emotions.

- Combine multiple modalities for reliable recognition.
- Calibrate models to individual users’ expression styles.
- Treat emotional misinterpretation as a learning opportunity.

Paper 3 — Schuller (2018): “Speech Emotion Recognition: Two Decades in a Nutshell”

Focus / Contribution

- Historical review of 20 years of speech-emotion research.
- Connects early paralinguistic work with modern deep-learning practices.

Key Insights

- Field shifted from handcrafted acoustic features to end-to-end learning.
- Human annotation inconsistency demands confidence or uncertainty modeling.
- Benchmark challenges (AVEC, Interspeech) drive progress.
- Gamified or crowdsourced labeling expands data diversity.- Projects such as *iHEARu-PLAY* demonstrate that gamification—turning data labeling into a fun or rewarding activity—attracts non-expert contributors and increases the cultural and emotional diversity of training data, which is vital for creating inclusive systems.

Design & Application Findings

- Show confidence levels in AI feedback (e.g., “70 % likely sad”).
- Encourage reflection rather than right/wrong evaluation.- Instead of scoring users as “correct” or “incorrect” in identifying emotions, the system should open space for reflection (“Does this match how you felt?”). This approach turns recognition into a conversation that builds emotional understanding and empathy.

Use gamified or narrative activities for emotional engagement.- Inspired by Schuller’s findings, emotional-training systems can employ story-based scenarios or game mechanics to make practice interactive and motivating. For example, users could earn progress by expressing or recognizing emotions within a story, turning learning into active play.

Paper 4 — Parada-Cabaleiro et al. (2020): “DEMoS – Italian Emotional Speech Corpus”

Focus / Contribution

- Developed an elicited (not acted) emotional-speech dataset using multiple mood-induction methods.

Key Insights

- Combining music, film, and empathy tasks produces authentic emotional speech.
- Self-assessments are unreliable; expert validation improves data quality.
- Emotion ambiguity (e.g., guilt, surprise) is natural and informative.
- Larger, more varied datasets outperform small curated ones.
- Emotional-awareness screening (alexithymia) affects data validity.

Design & Application Findings

- **Use safe narrative or music-based induction to elicit emotions.- Emotional-training environments can use storytelling, imagery, or background music to help users access authentic feelings safely—replicating DEMoS's elicitation strategies.**
- **Pair user self-ratings with AI or therapist-style feedback.**
- **Treat ambiguous emotions as valuable learning moments.**
- **Adapt interface for users with limited emotional awareness.- Some individuals may have difficulty expressing emotions vocally. Adaptive feedback, gentle prompting, or multimodal cues (text or visual aids) can make training more inclusive and effective.**

Paper 5 — Schuller et al. (2020): "INTERSPEECH Computational Paralinguistics Challenge"

Focus / Contribution

- **Benchmarks emotion recognition under real-world constraints: elderly voices, masked speech, and breathing sounds.**

Key Insights

- **Emotional cues change with age, masks, and respiration patterns.**
- **Deep CNN-LSTM models remain robust despite degraded input.**
- **Breathing and pauses convey emotional and physiological states.**
- **Ethical inclusion of vulnerable groups (elderly, clinical) is crucial.**

Design & Application Findings

- **Calibrate system sensitivity to each user's voice baseline.**
- **Incorporate breathing and rhythm data into emotional feedback.- Emotional-training interfaces can use breathing rate and speech pauses as part of feedback — for example, guiding users to notice tension through shallow breathing or calmness through steady rhythm.**
- **Adjust for low-energy or atypical expressions.- Users with limited vocal intensity or physical constraints (e.g., elderly, depressed, or fatigued speakers) may not exhibit strong emotional cues. Systems should avoid misclassifying these as "neutral" and instead provide sensitive, context-aware feedback.**
- **Ensure accessibility and inclusivity in emotion-training environments.**

Paper 6 — Radford et al. (2022): "Whisper – Robust Speech Recognition via Large-Scale Weak Supervision"

Focus / Contribution

- **Introduces a multilingual, multitask Transformer trained on 680 000 hours of weakly supervised audio–text data.**

Key Insights

- **Achieves near-human transcription and translation accuracy.**
- **Robust to noise, accents, and low-quality recordings.**
- **Performs multiple tasks: transcription, translation, language detection, and voice-activity recognition.**
- **Performance scales reliably with model size and data volume.**

Design & Application Findings

- **Provides reliable real-time transcription for emotional analysis.**
- **Enables multilingual emotional-training systems.**

- **Supports natural speech capture in interactive storytelling or therapy.**
- **Forms the technical backbone for robust, cross-lingual emotion recognition**

Conversational Interactions

Paper1. Conversational Agents in TherapeuticInterventions for NeurodevelopmentalDisorders: A Survey

Method: This paper synthesizes the literature on the development of conversational agents for skill intervention in individuals with Neurodevelopmental Disorders (NDD). It formulates a set of targeted research questions concerning the participants, the agents, and the methodological design of the studies to systematically analyze the current state of research in this field.
Contribution/Focus:

1. provide systematic synthesis of the state of art in the field along different dimensions
    a. related to participants
        i. 87% of studies focused on children (<18 years)
        ii. Gender distribution: 48% male with NDD, 9% female with NDD
        iii. 92% of studies focused on Autism Spectrum Disorder (ASD)
    b. related to agent
        i. 96% addressed communication and social skills broadly
        ii. 76% focused on training skills, 24% focused on assessment, 4% for both
        iii. interaction modalities mostly based on speech (both input and output) rather than gestures, gaze, tough, expressions, sounds….
        iv. 88% agents have no wake actions
        v. 79% Socially Assistive Robots (SARs) + 16% Embodied Conversational Agents (ECAs) + 8% Intelligent Personal Assistants (IPAs)
        vi. 75% humanoid+ 12% animal-like
        vii. Only 4% of agents had emotion recognition capabilities
        viii. 75% task-based interactions + 25% free interactions
    c. related to study
        i. Average duration: 60 days (SD = 79.66)
        ii. LCS algorithm effective for measuring joint attention
        iii. SARs and ECAs effective for emotional and social skills training; multi-modal interaction crucial; physical embodiment enhances engagement
2. provide recommendations for the future study+ create checklist(not shown here since not related to our project)
    a. design recommendations
        i. Embodiment: Prioritize socially assistive robots or embodied conversational agents; avoid disembodied forms
        ii. Shape: Start with simple shapes and gradually transition to human-like forms
        iii. Gender: Recommend using gender-neutral agents
        iv. Interaction Modalities: Provide multimodal interaction (speech, gesture, gaze, etc.) with modular enable/disable options
        v. Conversation Initiative: Design agents that proactively guide conversations
        vi. Wake Actions: Select appropriate wake actions based on user needs; develop turn-taking capabilities
        vii. Emotion Recognition: Enhance emotion recognition and social cue analysis capabilities
    b. Methodological recommendations
        i. Participants: Include caregivers; balance gender and age distribution
        ii. Disorder Scope: Expand beyond autism to include other NDD types
        iii. Study Design: Adopt randomized controlled trials and longitudinal studies; combine task-based and free interaction
        iv. Data Collection: Use mixed-methods approaches (quantitative + qualitative)

3. draw a research agenda that encompasses the directions for future work
    a. Priority: the most urgent tasks are not developing more agents with similar functions, but addressing foundational, bottleneck issues—namely, accessibility (democratization tools), suitability (user-centered design), and reliability (core technologies).
    b. challenges: develop personalized ASR for speech disorders, create modular architectures allowing therapist configuration.

Insight to our project: Consider designing a virtual narrator for our LLM storytelling system, incorporating multimodal interaction (voice, facial expressions, gestures) and emotion recognition capabilities where feasible.

Paper2. Empowering Conversational Agents using Semantic In-Context Learning

Contribution/focus: the LLMs are trained based on public datasets and may not be aware of private knowledge sources that could help them to resolve incoming inquiries more accurately. It's hard to apply fine-tuning LLMs due to high cost. frequent updates and privacy problems. The proposal of Semantic In-Context Learning (S-ICL) solved this problem.
The Semantic In-Context Learning (S-ICL) model

1. Semantic Search
    a. Vectorization: Convert your entire knowledge base and the user's current query into mathematical vectors (embeddings) using a model (e.g., Sentence-BERT). These vectors capture the semantic meaning of the text.
    b. Similarity Matching: The system calculates the similarity (e.g., using cosine similarity) between the user's query vector and all content vectors in the knowledge base.
    c. Retrieval: Return the top-K most similar snippets (e.g., the 3 most relevant Q&A pairs or document paragraphs).
2. In-Context Learning
    a. Construct the Prompt: Create a carefully designed prompt containing:
    b. Instruction: Tells the model what to do (e.g., "Please answer the question based on the provided examples.").
    c. Examples/Demonstrations: Insert the most relevant snippets retrieved in Step 1 as "examples" or "context" into the prompt.
    d. Query: The user's actual current question.
    e. Generate the Answer: Send this assembled prompt to the LLM (e.g., ChatGPT). The LLM, leveraging its powerful comprehension capabilities and the provided "examples," generates an answer that is both accurate and contextually appropriate.

Main findings:

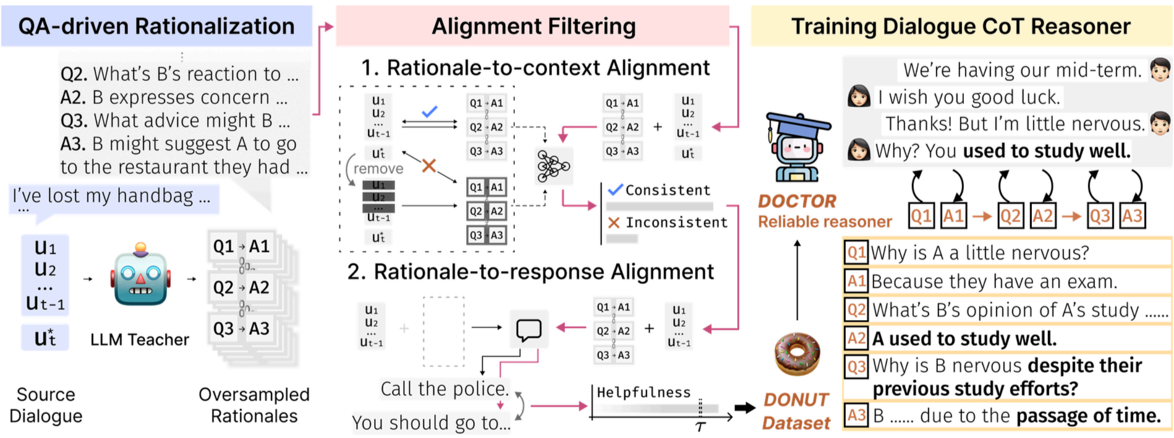
1. S-ICL delivers the best overall performance.
2. Prompt phrasing can control output style.
3. The best method depends on data characteristics.

Key insights to our project: Relying solely on LLMs to generate stories from scratch carries risks, such as producing inappropriate or therapeutically misaligned content. The optimal architecture integrates a therapist-curated story knowledge base, using semantic search to dynamically retrieve the most relevant narrative templates, which the LLM then adapts and personalizes.

Paper3. Dialogue Chain-of-Thought Distillation for Commonsense-aware Conversational Agents

Contribution/focus: This paper proposes a method for **constructing a high-quality Dialogue Chain-of-Thought (CoT) dataset and a dedicated reasoning model by distilling the reasoning capabilities of Large Language Models (LLMs)**. Experiments demonstrate that this method significantly enhances the commonsense reasoning ability of dialogue systems, resulting in more coherent, natural, and informative responses.

Explanation of this method:



1. QA-Driven Rationalization (Generating Candidates)
    a. Use ChatGPT to create multiple Q&A sequences
    b. Each sequence explains the dialogue step-by-step
    c. Questions follow commonsense templates (e.g., "What is the speaker's intent?")
    d. Answers reveal implicit information needed for response generation

2. Filter unreliable reasoning to get high-quality DONUT dataset
    a. Context Alignment: Train a classifier to distinguish between "factual chains" (based on full context) and "counterfactual chains" (based on truncated context).
    b. Response Alignment: Calculate the probability ratio to retains only reasoning chains that significantly improve the accuracy of response prediction.

3. Training the DOCTOR Model based on DONUT dataset.

Main findings:
1. Effective Distillation: The "generate-and-filter" framework successfully extracts high-quality reasoning chains from unreliable large language models.
2. DOCTOR Outperforms Teacher: The specialized 1.3B DOCTOR model surpasses its teacher model ChatGPT in dialogue reasoning tasks.
3. Quality Over Quantity: Small-scale, high-quality data filtered through alignment produces better models than large-scale noisy data.
4. Plug-and-Play Enhancement: DOCTOR's reasoning chains can directly improve response quality in any dialogue model.

Limitations:
1. Limited scope: Only tested on open-domain two-party dialogues, not covering task-oriented or multi-party scenarios.
2. Rigid reasoning: Uses fixed-step reasoning chains without dynamic adjustment capability.
3. Data risk: Training data is entirely machine-generated without human annotation verification.

Insights to our project: Drawing on the dialogue reasoning chain methodology from the paper, we can construct an emotion-specific reasoning process to form a multi-step cognitive training path of "emotion recognition → causal analysis → expression suggestion." Next, adopting the paper's "generate-and-filter" framework, we will develop emotional alignment filters to ensure the reasoning content aligns with the story's tone and maintains therapeutic effectiveness. Finally, based on the DOCTOR architecture, we will develop an Emo-DOCTOR model specifically designed to analyze emotional cues, generate cognitive reasoning chains, and provide personalized training guidance, thereby establishing a reliable emotional training environment.

Paper4. Reasoning LLMs for User-Aware Multimodal Conversational Agents

Background/problems:

1. The "Cold-Start" Problem: Traditional personalized systems fail during the initial interaction with a new user because no prior data (e.g., preferences, interaction history) exists.
2. Static User Modeling
3. Lack of Integrated Multimodal Reasoning

Method: The research designed and implemented a multi-stage, multimodal reasoning system based on a Retrieval-Augmented Generation (RAG) architecture, and conducted comprehensive validation through automated evaluation, human assessment, and ablation studies.
Contribution/focus: proposal of USER-LLM R1 model

1. Input: the user's facial image and their current textual query
2. User Identification & Retrieval:
    - The User Encoder processes the facial image and converts it into a mathematical representation (an embedding vector).
    - This vector is used to query a database to check if this user exists.
    - Path A (Cold Start): If the user is new, the User-VLM component is activated.
    - Path B (Returning User): If the user exists, the RAG (Retrieval-Augmented Generation) component retrieves the user's stored profile and past conversation history.
3. Initial Profile Generation (Cold Start):
    - The User-VLM analyzes the user's image and query to generate a textual initial user profile (e.g., "The person appears to be a Southeast Asian female, approximately 60-69 years old.").
4. Chain-of-Thought Reasoning & Response Generation:
    - The CoT Reasoning LLM receives all available information: the current query, the user profile (from User-VLM or the database), and the retrieved context (if any).
    - Instead of generating an answer immediately, it breaks down the task into intermediate reasoning steps. For example:
        - Step 1: "The user is an elderly person and might be unfamiliar with complex technical jargon."
        - Step 2: "Their question about scanning a document suggests they might be new to using smart devices."
        - Step 3: "Therefore, my response should avoid terms like 'OCR' or 'resolution' and instead use simple, step-by-step instructions like 'take a photo' and 'save to album'."
    - After this internal reasoning process, it produces two key outputs:
        - Output 1: An updated user profile (e.g., increasing confidence that the user is "not tech-savvy"), which is sent back to the database for future interactions.
        - Output 2: The final personalized response (a simple, step-by-step guide on how to scan a document).

Main findings: Through the deep integration of vision-language models with chain-of-thought reasoning large language models, the proposed USER-LLM R1 framework effectively addresses the cold-start problem, significantly outperforms existing methods in generating high-quality personalized responses, and achieves performance comparable to models with substantially larger parameter counts—where reasoning capability is proven to be the core driver of performance enhancement.

Insights to our project：Our project also faces the cold-start problem with new users. We will adopt the core concept from this research by implementing dynamic user profile updates, ensuring each interaction becomes progressively more aligned with the user's current emotional capabilities.

Paper5. Adapting Large Language Models for Education: Foundational Capabilities, Potentials, and Challenges

Goal: To systematically assess the core capabilities, potential, and challenges of Large Language Models in education, and to outline directions for building next-generation intelligent educational systems.

Method: A literature review analyzing the state-of-the-art and performance of LLMs across five key educational capabilities: Mathematics, Writing, Programming, Reasoning, and Knowledge-based QA.

Main findings:

1. Uneven Capabilities: No single LLM excels in all areas. GPT-4 is the strongest overall, but all models have specific weaknesses.
2. Progress with Problems:
    - Math: Good at solving complex problems using tools, but makes errors in basic arithmetic.
    - Writing: High quality in human evaluation, but standard automated metrics (e.g., ROUGE) are inadequate.
    - Programming: Strong at code generation, but weak at debugging and providing explanations.
    - Reasoning: Prompting techniques (e.g., Chain-of-Thought) help, but inherent reasoning skills remain limited.
    - Knowledge QA: Prone to "hallucination," requiring Retrieval-Augmented Generation (RAG) for mitigation.
3. Two Paths: A single, all-powerful unified model is currently infeasible; a Mixture-of-Experts (MoE) approach is more viable but architecturally complex.

Limitations in the field:

1. LLM flaws:
    - Hallucinations issues: may provide plausible but incorrect information
    - Weak complex reasoning: Performance significantly declines when handling complex problems
    - Limited specialized knowledge
2. Evaluation gaps:
    - Current metrics don't capture educational value

3. Practical barriers:
    - MoE framework is theoretical
    - Can't replace human teachers' insight and support

Insights to our project:

- Establish safety guardrail mechanisms to prevent harmful emotional advice
- Consider developing a therapist monitoring interface to allow professionals to supervise the training process
- Integrate retrieval-augmented generation (RAG) and incorporate psychological knowledge bases to ensure response accuracy

Conversational Agents in Software Engineering: Survey, Taxonomy and Challenges

— QUIM MOTGER and XAVIER FRANCH

— JORDI MARCO

CCS Concepts: · General and reference → Surveys and overviews;
Human-centered computing → Natural language interfaces.

1. Research Method

The study follows a rigorous tertiary study methodology (a systematic review of secondary studies) based on Kitchenham's guidelines.

- Data Sources: Major digital libraries like ACM, IEEE Xplore, Scopus, Web of Science, etc.
- Study Selection: An initial 136 studies were filtered down to a final set of 25 high-quality secondary studies using strict Inclusion/Exclusion Criteria and a Quality Assessment.
- Research Questions: The study addresses one General Question (GQ) and five Specific Questions (SQ):
    - GQ: What is the current state of research in the field of conversational agents?
    - SQ1: What research has been published? (Focus on Terminology, Domains, Goals)
    - SQ2: Which HCI features impact user experience?
    - SQ3: Which technical methods and technologies are used?
    - SQ4: Which methodological approaches are used for training, testing, and evaluation?

    - SQ5: What are the reported research challenges?

2. Key Findings & Proposed Taxonomies

SQ1: Published Research (Terminology, Domains, Goals)

- Terminology (F1): The field lacks a standardized taxonomy. The terms "Conversational Agent" and "Chatbot" are the most common and are predominantly used as synonyms.
- Domains (F2): CAs are applied across six major domains:
    1. Daily Life (e.g., smart-home assistants, tourism)
    2. Commerce (e.g., customer service, e-commerce)
    3. Business Support (e.g., internal process automation)
    4. Technical Infrastructure (e.g., software engineering assistance, IT support)
    5. Healthcare (e.g., therapy management, patient monitoring)
    6. Education (e.g., e-learning tools, automated tutoring)
- Goals (F3): The primary goals of integrating CAs are:
    1. User Support
    2. Information Request
    3. User Engagement
    4. Action Execution
    5. User Training
    6. Information Collection

SQ2: Impact of HCI Research (HCI Features)

- The study proposes a two-dimensional taxonomy of 69 HCI features (or "social cues"):
    - Dimension 1: Modality (from Feine et al.)
        - Verbal, Visual, Auditory, Invisible
    - Dimension 2: Footing (from Van Pinxteren et al.)
        - Human Similarity: How relatable the agent is to a human (user-independent).
        - Individual Similarity: How well the agent adapts to the individual user (user-dependent).
        - Responsiveness: How the agent reacts to a specific interaction.
- Key Insight: Research heavily focuses on verbal content for human similarity and responsiveness, but there is a significant gap in verbal content for individual similarity.

SQ3: Technical Methods & Technologies

- Design Dimensions (F5): Six key dimensions for classifying CAs:
    - Prescriptiveness: Task-Oriented vs. Non-Task-Oriented.
    - Knowledge Base: Open-Domain vs. Closed-Domain.
    - Service: Interpersonal (general) vs. Intrapersonal (personalized).
    - Response Generation: Deterministic (Rule-based) vs. AI-based (Retrieval or Generative).
    - Interaction: Text, Voice, or Multi-modal.
    - Human Aid: Autonomous vs. Human-Mediated.
- Technical Implementation (F6): Two major generations of CAs:
    - Deterministic (1st Gen): Rule-based (e.g., Pattern Matching with AIML). Mature, simple, but static.
    - AI-Based (2nd Gen): Uses Machine/Deep Learning (e.g., Intent Classifiers, RNNs, LSTMs, Seq2Seq). Adaptive, dynamic, but complex and data-hungry.
    - Hybrid approaches that combine both are promising.
- Context Integration (F7): Identified as a key future area. Strategies are categorized by:
    - Purpose (e.g., improve engagement)
    - Source (e.g., user profile)
    - Mechanism (e.g., adaptive knowledge base)
    - Object (e.g., agent's response)

SQ4: Training, Testing & Evaluation

- Data-sets (F8): A significant research gap. There is a lack of structured discussion on data sources, which are often domain-specific (e.g., medical repositories, chat logs).
- Quality & Evaluation Methods (F9):
    - Quality Characteristics are mapped to the ISO/IEC 25010 standard, with emphasis on:
        - Functional Suitability (Effectiveness, Appropriateness)
        - Usability (User Satisfaction, Learnability)
        - Performance Efficiency
        - Security (Confidentiality, Integrity)
    - Evaluation Methods:
        - Qualitative: Interviews, Questionnaires (using Likert scales).
        - Quantitative: Dialogue Tracking (using metrics like accuracy, F-measure), Surveys.

SQ5: Future Research & Challenges

Six major challenges were identified to guide future work:

1. Adoption in Domain/Target-Specific Scenarios: Moving from proof-of-concepts to real-world tools, especially for specific user groups (e.g., elderly).
2. Improving User Engagement via Perceived Quality: Enhancing usability, satisfaction, and trust to keep users engaged.
3. Extending HCI Features for Individual Similarity: Personalizing verbal content and non-verbal features (e.g., appearance, speech) based on individual users.
4. Achieving Full Personalization: Leveraging user profiling and context-awareness for truly adaptive agents.
5. Expanding Knowledge on Data: Creating taxonomies of open-source data repositories to support AI-based CA development.
6. Extending Qualitative Evaluation Methodologies: Better methods to capture and measure user-perceived quality and appropriateness.

3. Limitations

- Construct Validity: Risk of missing relevant studies was mitigated by a comprehensive search and snowballing.
- Conclusion Validity: Personal bias was reduced through collaborative author sessions. The provided replication package ensures transparency and replicability.

4. Conclusion

This survey provides a holistic overview of the conversational agents field. It demonstrates the field's interdisciplinary nature and the convergence of techniques from HCI, Software Engineering, and AI. While deterministic agents are mature, the future lies in adaptive, AI-based, and fully personalized agents. The proposed taxonomies and identified challenges are intended to help researchers classify their work and lay the groundwork for future advancements in natural language interfaces.

Integrating Large Language Models into Therapeutic Education for Children with Dyslexia: a Multimodal Framework

Francesco Piferi , Giovanni Caleffi , Giulia Valcamonica , Pietro Crovari and Franca Garzotto

Focus / Contribution

Proposes the ChatCare framework, a multimodal system for therapeutic education that deeply integrates Large Language Models (LLMs) with real-time emotion recognition and event-driven proactive interaction to create a personalized and highly engaging learning environment for children with dyslexia.

Key Insights

1. LLMs enable contextualized learning support: The LLM provides targeted guidance and encouragement based on the user's current learning activity (e.g., a spelling exercise).
2. Emotion is a critical dimension in child-centric interfaces: Analyzing emotion from text and voice allows the system to dynamically adapt the chatbot's tone and strategy for empathetic feedback.
3. Proactive interaction significantly boosts engagement: The system initiates conversations based on user behavior (e.g., repeated failures, task completion), functioning as an active learning partner rather than a passive respondent.
4. Multimodal interfaces bridge functionality and user experience: A cartoonish aesthetic and an animated avatar translate complex backend technology into a friendly and engaging interaction.

Design & Application Findings

- Encode expert knowledge into system prompts: Embed therapeutic strategies and educational logic into the LLM through carefully engineered prompts.
- Utilize an event-driven mechanism for proactive care: Trigger chatbot interventions for help and encouragement based on detected events like consecutive failures or session completion.
- Calibrate emotion recognition for children's expression: Systems must be tuned to identify atypical or subtle emotional cues in children to avoid misclassification as "neutral."
- Construct a unified environment integrating learning and emotional support: Co-locate the activity and conversational interfaces to seamlessly blend academic tasks with emotional support.

"It’s a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents

Zhiping Zhang, Bingsheng Yao, Michaele Jia, Hao-Ping (Hank) Lee, Sauvik Das, Ada Lerner, Dakuo Wang, Tianshi Li
Northeastern University, Boston, MA, USA
Rensselaer Polytechnic Institute, Troy, NY, USA
Carnegie Mellon University, Pittsburgh, PA, USA

Focus / Contribution
Proposes a user-centered empirical framework that, through the analysis of real-world ChatGPT conversation logs and semi-structured interviews, systematically uncovers privacy disclosure behaviors, flawed mental models, and systemic design pitfalls when users interact with LLM-based conversational agents, highlighting the asymmetric trade-offs between utility, convenience, and privacy.

Key Insights

- Users face asymmetric trade-offs between privacy and utility: Users often believe they must sacrifice privacy to access ChatGPT's powerful capabilities, viewing it as a "fair game," despite significant inaccuracies in their mental models.
- Flawed mental models undermine privacy decision-making: Most users hold misconceptions about how LLMs generate responses and how their data is used for training, impairing their ability to accurately assess privacy risks.
- Human-like interactions encourage sensitive disclosures: ChatGPT's anthropomorphic conversational style leads users to progressively reveal more personal or others' information, creating cascading privacy risks.
- Dark patterns in system design hinder privacy-protective behaviors: ChatGPT's privacy settings incorporate dark patterns such as bundled options and hard-to-find opt-out forms, limiting users' effective control over their data

Design & Application Findings

- Design privacy support aligned with user mental models: Systems should proactively address user misconceptions about LLM functionality or align interface design with user expectations.
- Develop assistive privacy-enhancing tools: Automated or semi-automated tools should be designed to help users identify and sanitize sensitive information in their inputs, reducing manual effort.
- Exercise caution in using user data for model training: If user data is used for training, memorization risks must be clearly communicated, with easily discoverable and granular opt-out controls.
- Prioritize local models where feasible: For specific use cases, lighter models that run locally should be considered to fundamentally avoid data transmission and memorization risks.

Proactive Conversational Agents in the Post-ChatGPT World

Lizi Liao[1], Grace Hui Yang[2], Chirag Shah[3]
[1]Singapore Management University, Singapore
[2]Georgetown University, Washington, DC, USA
[3]University of Washington, Seattle, WA, USA

Focus/Contribution
This tutorial provides a comprehensive review of methods for developing proactive conversational agents that can initiate conversations, shift topics, and offer context-aware recommendations, addressing the critical limitation of passivity in current LLM-based systems like ChatGPT.

Key Insights

- Proactivity represents the next frontier for conversational AI: Current LLM-based agents remain predominantly reactive, limiting their effectiveness in scenarios requiring active engagement such as exploratory search and complex decision-making.
- Multiple technical approaches enable conversational proactivity: Strategic methods including learning-to-ask algorithms, topic shifting mechanisms, and reinforcement learning with subgoals provide structured pathways for agents to lead conversations.
- Quality control becomes crucial with increased agent autonomy: As agents take more conversational initiative, ensuring safety, mitigating hallucinations, and maintaining alignment with human values emerge as paramount concerns.
- Proactive agents demand novel evaluation frameworks: Traditional metrics prove inadequate for assessing how effectively agents drive conversations toward goals while maintaining user engagement and trust.

Design & Application Findings

- Architect conversational strategies into system design: Implement techniques such as mixed-initiative dialogue, target-guided topic shifting, and reward-augmented reinforcement learning to embed proactive reasoning capabilities.
- Implement multi-layered quality control mechanisms: Integrate real-time safeguards including language detoxification, conservative RL policies, and hallucination detection to manage risks associated with agent-led dialogue.

- Design for interactive and adaptive evaluation: Incorporate human-in-the-loop testing and iterative feedback systems to assess and refine proactive agent behaviors in real-world scenarios.
- Balance autonomy with user control: Ensure agents can guide conversations while preserving user ability to override, correct, or redirect dialogue to maintain trust and usability.

Talk2Care: An LLM-based Voice Assistant for Communication between Healthcare Providers and Older Adults

Focus / Contribution

The system features a dual-module architecture: a patient-facing voice assistant for intuitive health information collection, and a provider-facing dashboard for automated conversation summarization and key information highlighting. By leveraging the natural language capabilities of LLMs while deliberately avoiding the provision of specific medical advice, the system focuses on enhancing communication efficiency, information richness, and workflow support without replacing clinical judgment.

Key Insights

- LLMs enable structured, protocol-driven data collection: Through carefully engineered prompts, the system conducts natural, multi-turn conversations with older adults, gathering detailed symptom descriptions and health concerns based on customizable clinical protocols.
- Voice interface overcomes accessibility barriers: The voice-based interaction reduces the digital literacy burden for older adults, providing a more accessible and familiar modality compared to text-based portals.
- Automated summarization and prioritization alleviate provider workload: The dashboard condenses patient-VA conversations into structured clinical notes, highlights critical information, and suggests risk levels, enabling providers to triage and respond more efficiently.
- AI-mediated communication reduces psychological barriers: Older adults reported feeling less judged and more comfortable sharing information with the VA, while providers appreciated the reduction in repetitive communication tasks and emotional labor.

Design & Application Findings

- Dual-module architecture supports closed-loop communication: The voice assistant collects information from older adults, and the dashboard presents processed insights to providers, creating an integrated asynchronous communication pipeline.
- Prompt engineering embeds clinical knowledge and ethical safeguards: Multi

GENERATION WITH LLMs

1) Supporting GenAI-driven Creation of Images for Educational Storytelling — *Valcamonica & Garzotto*

Goal: Introduces Descriptor, a ChatGPT-based assistant that turns scene sentences into structured, detailed prompts for text→image engines to improve visual coherence across multi-image stories for children.

Methods: design of prompt-enrichment templates (character/object/action/environment/style) and a small qualitative case study comparing images generated with vs without Descriptor using DALL·E and similar models. Experts evaluated images for consistency and clarity.

Main findings: Descriptor-generated prompts produced more consistent character appearance, style coherence, and contextual relevance across scenes; users (educators) found it quick to learn and productive.

Limitations: evaluation is preliminary/qualitative; single default art style (children's-book) reduces flexibility; method explored mainly in educational storytelling (not yet tested for emotionally sensitive therapeutic material).

Relevance to our project:

- **Directly applicable for producing visually consistent scenes in interactive storytelling for emotional training (important so emotional cues remain stable across episodes).**

- **Descriptor's structured descriptors are useful input to any T2I model and can be extended to encode emotion labels, facial expression attributes, and pacing cues for therapeutic stories.**

2) SS-GEN: A Social Story Generation Framework with Large Language Models — *Feng et al.*

Goal: Creates an LLM-based pipeline (STARSOW) to generate high-quality Social Stories (used for ASD interventions) under strict constraints (structural clarity, descriptive orientation, situational safety). Builds a filtered dataset (≈5K stories) and fine-tunes smaller models on it.

Methods: hierarchical, constraint-driven prompting to expand seed chapters → many titles → story content; human filtering + quality assessment checklist; experiments fine-tuning small models and human/GPT evaluation.

Main findings: STARSOW + human filtering produces diverse, constraint-adherent Social Stories at scale; fine-tuned smaller models can match larger LLM behaviour for the task at lower cost; generated stories score well on safety and structure metrics.

Limitations: Human filtering still required (≈25–30% outputs filtered); the approach focuses on textual social stories (less on multimodal / images); may need careful domain adaptation for different emotional conditions.

Relevance to our project:

- **High relevance for story generation component: STARSOW procedure and quality checklist give a concrete method to generate safe, structured narratives for emotional training (e.g., practice recognizing/expressing feelings).**
- **The dataset+fine-tuning approach is a practical path for obtaining reliable, low-cost LLMs that produce therapy-friendly texts.**

3) Playing repeated games with Large Language Models — *Akata et al.*

Goal: Uses behavioral game theory to study LLMs' behavior in iterated social interactions (2×2 repeated games), revealing their tendencies in cooperation, retaliation, coordination, and how prompting modifies behavior.

Methods: turn payoff matrices into prompts and run iterated games (10 rounds) between GPT-3 / GPT-3.5 / GPT-4 and scripted agents; analyze patterns across game families (Prisoner's Dilemma, coordination/Battle of the Sexes, win-win, etc.). Explore prompt interventions (ask LLM to predict opponent first, or tell it others can make mistakes).

Main findings:

- **LLMs do well when self-interest aligns with the objective (e.g., PD family) but struggle with coordination requiring conventions (e.g., alternation).**

- **GPT-4 is often unforgiving (retaliates after a single defection) but can be nudged to cooperate/coordinate by prompting it to predict others' actions or reminding it that others make mistakes (i.e., social chain-of-thought helps).**

Limitations: focus on simple 2×2 games (controlled but limited); pattern discovery is exploratory/post-hoc; behavior may depend on prompt framing and model version.

Relevance to your project:

- **Crucial insight: LLMs' social reasoning and willingness to adapt can be changed via prompting (ask it to predict or reason about the user's state first). That's directly useful for building empathetic, adaptive storytelling agents (improve Theory-of-Mind-like behaviour).**

- **Warns you that out-of-the-box LLMs may default to self-interested or rigid behaviours — so you'll need explicit scaffolds and controlled prompts to ensure supportive, forgiving, therapeutic responses.**

Week 2- user research based on emotional computing paper evaluation

User personas??? Includes user profiles, preferences, expertise, accessibility needs, and behavioral patterns.

USER- The primary users of this system are individuals with emotional or affective disorders who require structured yet engaging ways to recognize, express, and regulate their emotions. The interface therefore must provide a supportive and adaptive environment where users can explore emotions through interactive storytelling and conversational experiences. It should recognize vocal, linguistic, and behavioural emotional cues, offering real-time, empathetic feedback that encourages reflection rather than right/ wrong. The system should support diverse expression styles, cultural backgrounds, and communication abilities. The platform must allow personalization, adapting difficulty, tone, and emotional feedback based on each user's progress and emotional sensitivity. Secondary users require tools for monitoring emotional engagement and progress while ensuring data privacy and ethical handling of emotional information. Overall, the system should support emotional growth, self-awareness, and resilience through safe, adaptive, and human-centered interaction.

NAVIGATION- Navigation within the emotional training interface should be simple and goal-oriented, guiding users smoothly through each stage of emotional exploration. A clear, consistent layout will help users access with minimal effort as they can maintain orientation and reinforce a predictable mental model, reducing cognitive load for users who may already experience emotional stress. Search and filtering options allow quick access to specific stories or emotion themes, while guided interaction eg. prompts, dialogue choices, and feedback loops would ensure users always know what to do next. Overall, navigation should reinforce emotional safety while helping users reach their learning goals confidently and without confusion.

GOAL- The primary goal for users of the emotional training interface is to enhance their emotional understanding and regulation skills through engaging, interactive experiences. Users aim to recognize and express emotions more accurately, interpret emotional cues in others, and build confidence in managing emotional responses in real-life situations. The interface should help them reach these goals efficiently and

comfortably by providing structured storytelling exercises, real-time emotional feedback, and progress reflections. Every design element from dialogue prompts to visual feedback should support this goal by keeping users emotionally engaged, reducing confusion, and reinforcing learning through positive, empathetic interaction.

github

Reviewed Papers

Emotional Computing

Paper 1 — Catania & Garzotto (2020): “Emozionalmente – Crowdsourced Emotional Speech Corpus”

Focus / Contribution

- Created an Italian emotional-speech dataset using a web-based crowdsourcing platform.
- Evaluated human and AI emotion recognition performance (using Wav2Vec 2.0).

Key Insights

- Crowdsourced emotional data can be accurate and scalable.
- Deep-learning models outperform human listeners in emotion detection.
- Cultural and linguistic specificity greatly affect emotion recognition accuracy.
- Some emotions (fear, disgust) remain difficult to classify.

Design & Application Findings

- Enables automated, real-time speech-emotion feedback.
- Systems should tolerate ambiguity and multiple interpretations. Given that even humans misclassify emotions, design should not treat the AI’s reading as absolute truth.
 Instead, feedback could express confidence levels (“You sound slightly anxious, confidence 70%”) or offer interpretive reflection (“You may be expressing tension—does that feel right to you?”).
- Crowdsourcing or user-generated emotion samples can personalize training.

Paper 2 — Wani et al. (2021): “A Comprehensive Review of Speech Emotion Recognition Systems”

Focus / Contribution

- Reviews the full SER pipeline from data collection to classification.
- Compares traditional machine learning with modern deep learning methods.

Key Insights

- Deep models (CNN, RNN, LSTM) outperform traditional approaches.

- Dimensional emotion models (valence–arousal) better capture nuance.
- Multimodal fusion (speech + text + facial) increases robustness.
- SER remains limited by labeling bias, noise, and cross-cultural variation.

Design & Application Findings

- Use dimensional feedback rather than fixed emotion labels. Instead of telling users "You sound angry," the system could show a position on a valence–arousal map ("high arousal, negative tone"), helping users explore the gradient of their emotions.
- Combine multiple modalities for reliable recognition.
- Calibrate models to individual users' expression styles.
- Treat emotional misinterpretation as a learning opportunity.

Paper 3 — Schuller (2018): "Speech Emotion Recognition: Two Decades in a Nutshell"

Focus / Contribution

- Historical review of 20 years of speech-emotion research.
- Connects early paralinguistic work with modern deep-learning practices.

Key Insights

- Field shifted from handcrafted acoustic features to end-to-end learning.
- Human annotation inconsistency demands confidence or uncertainty modeling.
- Benchmark challenges (AVEC, Interspeech) drive progress.
- Gamified or crowdsourced labeling expands data diversity. Projects such as iHEARu-PLAY demonstrate that gamification—turning data labeling into a fun or rewarding activity—attracts non-expert contributors and increases the cultural and emotional diversity of training data, which is vital for creating inclusive systems.

Design & Application Findings

- Show confidence levels in AI feedback (e.g., "70% likely sad").
- Encourage reflection rather than right/wrong evaluation. Instead of scoring users as "correct" or "incorrect" in identifying emotions, the system should open space for reflection ("Does this match how you felt?").
- Use gamified or narrative activities for emotional engagement. Inspired by Schuller's findings, emotional-training systems can employ story-based scenarios or game mechanics to make

practice interactive and motivating.

Paper 4 — Parada-Cabaleiro et al. (2020): “DEMoS – Italian Emotional Speech Corpus”

Focus / Contribution

- **Developed an elicited (not acted) emotional-speech dataset using multiple mood-induction methods.**

Key Insights

- **Combining music, film, and empathy tasks produces authentic emotional speech.**
- **Self-assessments are unreliable; expert validation improves data quality.**
- **Emotion ambiguity (e.g., guilt, surprise) is natural and informative.**
- **Larger, more varied datasets outperform small curated ones.**
- **Emotional-awareness screening (alexithymia) affects data validity.**

Design & Application Findings

- **Use safe narrative or music-based induction to elicit emotions. Emotional-training environments can use storytelling, imagery, or background music to help users access authentic feelings safely—replicating DEMoS’s elicitation strategies.**
- **Pair user self-ratings with AI or therapist-style feedback.**
- **Treat ambiguous emotions as valuable learning moments.**
- **Adapt interface for users with limited emotional awareness. Some individuals may have difficulty expressing emotions vocally. Adaptive feedback, gentle prompting, or multimodal cues (text or visual aids) can make training more inclusive and effective.**

Paper 5 — Schuller et al. (2020): “INTERSPEECH Computational Paralinguistics Challenge”

Focus / Contribution

- **Benchmarks emotion recognition under real-world constraints: elderly voices, masked speech, and breathing sounds.**

Key Insights

- Emotional cues change with age, masks, and respiration patterns.
- Deep CNN-LSTM models remain robust despite degraded input.
- Breathing and pauses convey emotional and physiological states.
- Ethical inclusion of vulnerable groups (elderly, clinical) is crucial.

Design & Application Findings

- Calibrate system sensitivity to each user's voice baseline.
- Incorporate breathing and rhythm data into emotional feedback. Emotional-training interfaces can use breathing rate and speech pauses as part of feedback—for example, guiding users to notice tension through shallow breathing or calmness through steady rhythm.
- Adjust for low-energy or atypical expressions. Users with limited vocal intensity or physical constraints (e.g., elderly, depressed, or fatigued speakers) may not exhibit strong emotional cues. Systems should avoid misclassifying these as "neutral" and instead provide sensitive, context-aware feedback.
- Ensure accessibility and inclusivity in emotion-training environments.

Paper 6 — Radford et al. (2022): "Whisper – Robust Speech Recognition via Large-Scale Weak Supervision"

Focus / Contribution

- Introduces a multilingual, multitask Transformer trained on 680,000 hours of weakly supervised audio–text data.

Key Insights

- Achieves near-human transcription and translation accuracy.
- Robust to noise, accents, and low-quality recordings.
- Performs multiple tasks: transcription, translation, language detection, and voice-activity recognition.
- Performance scales reliably with model size and data volume.

Design & Application Findings

- Provides reliable real-time transcription for emotional analysis.
- Enables multilingual emotional-training systems.
- Supports natural speech capture in interactive storytelling or therapy.
- Forms the technical backbone for robust, cross-lingual emotion recognition.

Conversational Interactions

Paper 1 — Conversational Agents in Therapeutic Interventions for Neurodevelopmental Disorders: A Survey

Method:
Synthesizes literature on conversational agents (CAs) for interventions in individuals with Neurodevelopmental Disorders (NDDs).

Key findings:

- 87% of studies focused on children (<18 years).
- 92% targeted Autism Spectrum Disorder (ASD).
- 79% of agents were socially assistive robots; only 4% had emotion recognition.
- Multimodal interaction and physical embodiment improve engagement.

Insights to Project:
Design a virtual narrator for the LLM storytelling system with multimodal interaction (voice, facial expressions, gestures) and emotion recognition where feasible.

Paper 2 — Empowering Conversational Agents using Semantic In-Context Learning

Focus / Contribution:
Proposes Semantic In-Context Learning (S-ICL) to enable context-aware dialogue without costly fine-tuning.

Key Insights:

- Uses semantic vector retrieval and context construction to answer queries accurately.
- S-ICL delivers high performance with flexible prompting.

Insights to Project:
Integrate a therapist-curated story knowledge base; use semantic retrieval to personalize and ensure therapeutic alignment.

Paper 3 — Dialogue Chain-of-Thought Distillation for Commonsense-Aware Conversational Agents

Focus / Contribution:
Proposes a dialogue reasoning model (DOCTOR) that distills LLM reasoning into coherent, commonsense-aware dialogues.

Key Insights:

- **Generate high-quality reasoning chains from LLMs (“generate-and-filter” method).**
- **Smaller distilled model outperforms its teacher in dialogue reasoning.**

Insights to Project:
Create an emotion-specific reasoning process for training: *emotion recognition → causal analysis → expression suggestion*. Develop an Emo-DOCTOR architecture for structured emotional reasoning.

Paper 4 — Reasoning LLMs for User-Aware Multimodal Conversational Agents

Focus / Contribution:
Proposes the USER-LLM R1 model integrating facial recognition and chain-of-thought reasoning for adaptive dialogue.

Key Insights:

- **Addresses the cold-start problem with new users.**
- **Dynamically updates profiles using multimodal inputs (facial image + query).**
- **Outperforms larger models in personalization and reasoning quality.**

Insights to Project:
Adopt dynamic user profiling to adapt emotional storytelling as users progress.

Paper 5 — Adapting Large Language Models for Education

Focus / Contribution:
Assesses LLMs’ educational capabilities in reasoning, writing, and knowledge QA.

Key Insights:

- **GPT-4 is strong but not universal.**
- **Hallucinations and limited reasoning persist.**
- **Retrieval-Augmented Generation (RAG) reduces inaccuracies.**

Insights to Project:
Add safety guardrails and therapist supervision. Integrate psychological knowledge bases for accuracy and safety.

Paper 6 — Conversational Agents in Software Engineering: Survey, Taxonomy and Challenges

Focus / Contribution:
Systematic tertiary review following Kitchenham's methodology.

Key Insights:

- **Defines CA taxonomies across domains, modalities, and goals.**
- **Identifies need for personalization, context-awareness, and improved HCI evaluation.**

Insights to Project:
Highlights the importance of personalized, human-centered conversational systems for emotional training.

Paper 7 — Integrating Large Language Models into Therapeutic Education for Children with Dyslexia

Focus / Contribution:
Proposes *ChatCare*, a multimodal LLM-based educational assistant with emotion recognition and proactive dialogue.

Key Insights:

- **Contextualized learning support through emotional cues.**
- **Emotion analysis allows adaptive feedback.**
- **Event-driven proactive interaction enhances engagement.**

Design Findings:

- **Embed expert knowledge into prompts.**
- **Trigger proactive feedback on key events (e.g., repeated errors).**
- **Integrate learning and emotional support seamlessly.**

Paper 8 — "It's a Fair Game": Privacy and Disclosure Risks with LLM Agents

Focus / Contribution:
Examines privacy behaviors and risks through ChatGPT interaction analysis and user interviews.

Key Insights:

- **Users trade privacy for convenience.**
- **Anthropomorphic design encourages oversharing.**
- **Dark patterns obscure data control.**

Design Findings:

- **Align privacy design with user mental models.**
- **Offer assistive tools to flag sensitive inputs.**
- **Favor local or hybrid processing for privacy.**

Paper 9 — Proactive Conversational Agents in the Post-ChatGPT World

Focus / Contribution:
Tutorial reviewing methods for proactive conversational AI.

Key Insights:

- **Proactivity improves engagement but increases risk.**
- **Reinforcement learning and topic-shifting enable initiative-taking.**
- **Requires human-in-the-loop testing and safeguards.**

Design Findings:

- **Implement proactive yet controllable dialogue.**
- **Integrate hallucination detection and safe reward mechanisms.**

Paper 10 — Talk2Care: An LLM-Based Voice Assistant for Healthcare Providers and Older Adults

Focus / Contribution:
Dual-module voice system improving patient–provider communication.

Key Insights:

- **Voice interface reduces accessibility barriers.**
- **Summarization aids providers.**
- **Patients disclose more comfortably to AI than humans.**

Design Findings:

- **Dual architecture for feedback loop.**
- **Embed ethical guardrails via prompt design.**

Generation with LLMs

1) Supporting GenAI-driven Creation of Images for Educational Storytelling — Valcamonica & Garzotto

Goal:
Introduces *Descriptor*, a ChatGPT-based assistant that turns scene sentences into structured prompts for text-to-image models, ensuring visual consistency.

Findings:

- **Improves character coherence and visual clarity.**
- **Easy for educators to use.**
- **Evaluations are qualitative and limited to educational settings.**

Relevance:
Useful for producing consistent emotional visuals in interactive storytelling.

---

2) SS-GEN: A Social Story Generation Framework — Feng et al.

Goal:
Builds *STARSOW*, an LLM pipeline for generating structured, safe social stories used in ASD interventions.

Findings:

- **Produces diverse, safe narratives.**
- **Human filtering remains necessary (~25–30%).**
- **Adapts well to emotionally sensitive applications.**
  Relevance:

  Provides a framework for structured, therapy-friendly emotional stories.

---

3) Playing Repeated Games with Large Language Models — Akata et al.

Goal:
Uses behavioral game theory to examine LLM cooperation and reasoning.

Findings:

- **GPT-4 can cooperate or retaliate depending on prompt framing.**
- **Predictive prompting improves coordination.**

Relevance:
Guides the design of empathetic, cooperative storytelling agents.

Current Products for LLM-Based Emotional Training

Several cutting-edge products address emotion recognition and expression training for people with emotional disorders:

AI Chatbots & Storytelling

Woebot and Wysa lead mental health chatbot solutions, using CBT principles for emotional support and reducing depressive symptoms by 64%. CuentosIE teaches emotional intelligence through interactive storytelling with 30 emotion taxonomies. StoryMate uses GPT-4 for personalized story-reading with adaptive, child-centered interactions.

Mobile Apps

EmoTEA helps children with autism (ages 6-12) recognize six basic emotions through gamified activities. InnerVoice combines AAC with AI-powered visual language learning, using 3D avatars and facial animation to improve social communication.

Virtual Reality

Floreo VR provides evidence-based social skills training through immersive scenarios (street safety, social interactions), with studies showing skill maintenance after use. VR interventions demonstrate effectiveness almost three-fourths of a standard deviation above traditional methods.

Social Robots

Milo (RoboKind) delivers comprehensive social-emotional curricula to 600+ students with 90% skill generalization rates. NAO and Pepper robots offer AI-generated educational content with emotion recognition capabilities. These robots provide non-judgmental, repeatable interactions particularly effective for autism therapy.

Wearable Tech

Superpower Glass uses augmented reality with Affectiva's emotion AI to provide real-time feedback on eight emotions, improving eye contact and emotional vocabulary in children with autism.

FDA-Approved Tools

Cognoa is the first FDA-authorized AI diagnostic tool for autism (ages 18 months-5 years), achieving 81% accuracy for positive diagnoses and addressing critical diagnostic delays.

Market Growth

The emotion AI market reached $2.9 billion in 2024 and projects to $16.97 billion by 2033. Digital therapeutics for ADHD/autism grew from $1.3 billion (2025) toward $5.4 billion by 2032, driven by increasing accessibility and evidence-based effectiveness.