Training Generative AI to support chemistry learning and with remediation specialists.

Sumana Chilakamarri, Navya Ammayath, Guy Toles, Dawn Evans, Anand Nagpurkar, Maureen Linden, Piyush Jha, Vijay Ganesh, Abigale Stangl.

Georgia Tech

Center for Inclusive Design and Innovation

Improve the human condition through equal access to technology-based and research-driven information, services, and products for individuals with disabilities.

Multisense.Studio

An interdisciplinary research group started by Dr. Abigale Stangl at Georgia Tech focused on the ethical design and development of inclusive services, products, and experiences through accessibility innovations and personalization.

Reasoning & Learning Group

Advances theory and practice of SAT/SMT solvers involve developing and applying logic-based methods to determine the satisfiability of complex computational and mathematical formulas, driving advances across software engineering, security, artificial intelligence, mathematics, and physics.
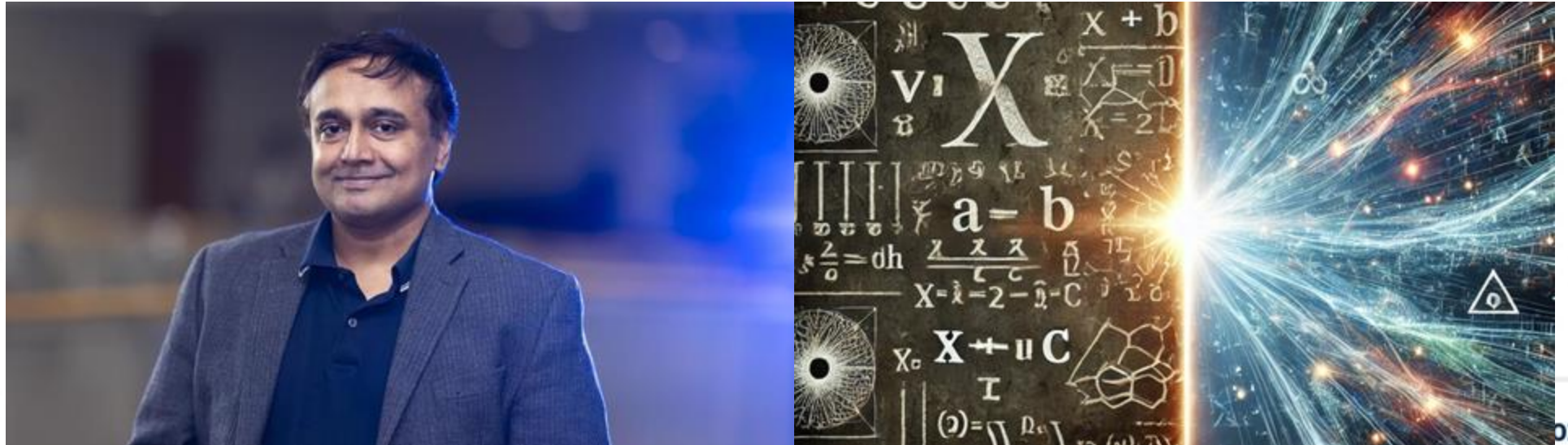
IDEaS & Microsoft

Advance GenAI research by supporting broad, interdisciplinary projects that explore both foundational innovations and novel applications across fields—while emphasizing responsible and ethical AI development to address key challenges and opportunities in the rapidly evolving landscape.

Chemistry

Improve the human condition through equal access to technology-based and research-driven information, services, and products for individuals with disabilities.

Bryan F. Shaw, Ph.D., Baylor
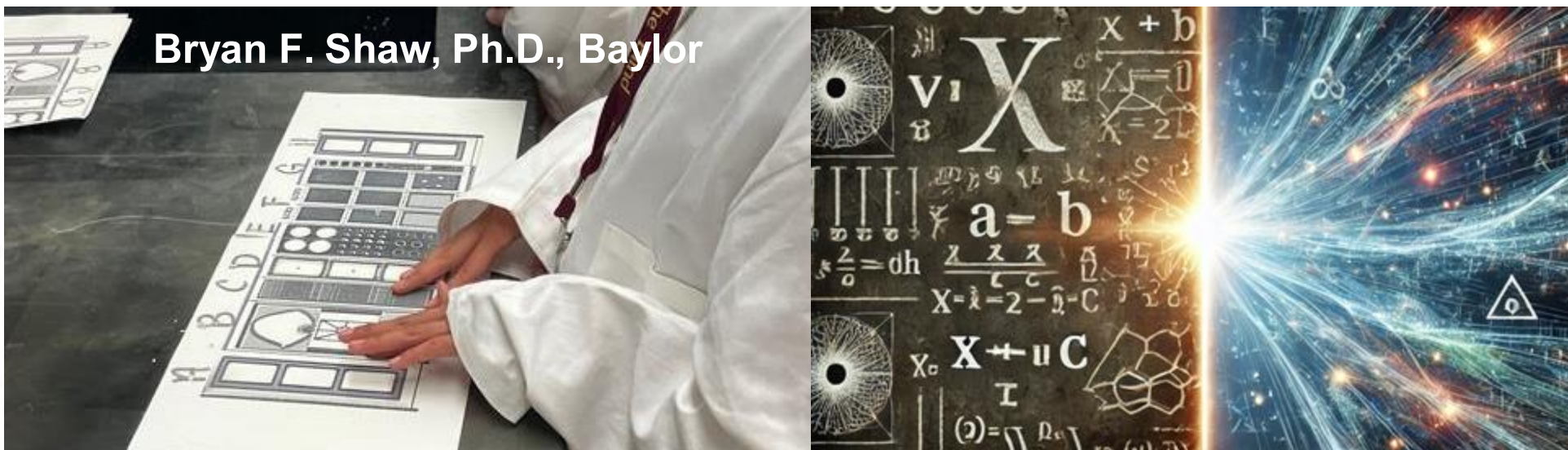
Problem Formulation

STEM materials need to be remediated to ensure that blind individuals or other multimodal learners can learn chemistry.

Bryan F. Shaw, Ph.D., Baylor

Hoby Wedler, Ph.D.

Georgia Tech
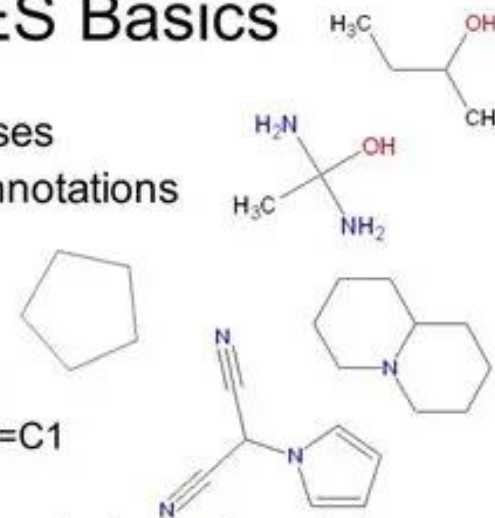
Problem Formulation

Chemistry notation—formulas, reactions, and molecular diagrams—creates major accessibility barriers for (BLV) learners.

Gen AI offers efficiency but often produces plausible yet wrong outputs, missing cues like stereochemistry, catalysts, or reaction conditions.

AI lacks true chemical understanding, so outputs must be grounded in chemical truth to ensure accessibility and trust.

SMILES Basics

- Branching: Parentheses
- Cycles: Numerical annotations
    - CCC(O)C
    - CC(N)(N)O
    - C1CCCC1
    - N12CCCCC1CCCC2
    - N#CC(C#N)N1C=CC=C1
- Extensions for
    - Inorganic atoms, unusual valence, formal charges, stereochemistry, aromaticity, reactions, etc.

Problem Formulation

Initial prompting experiments found that general AI models did not provide adequate alt-text for remediation specialists on our team.
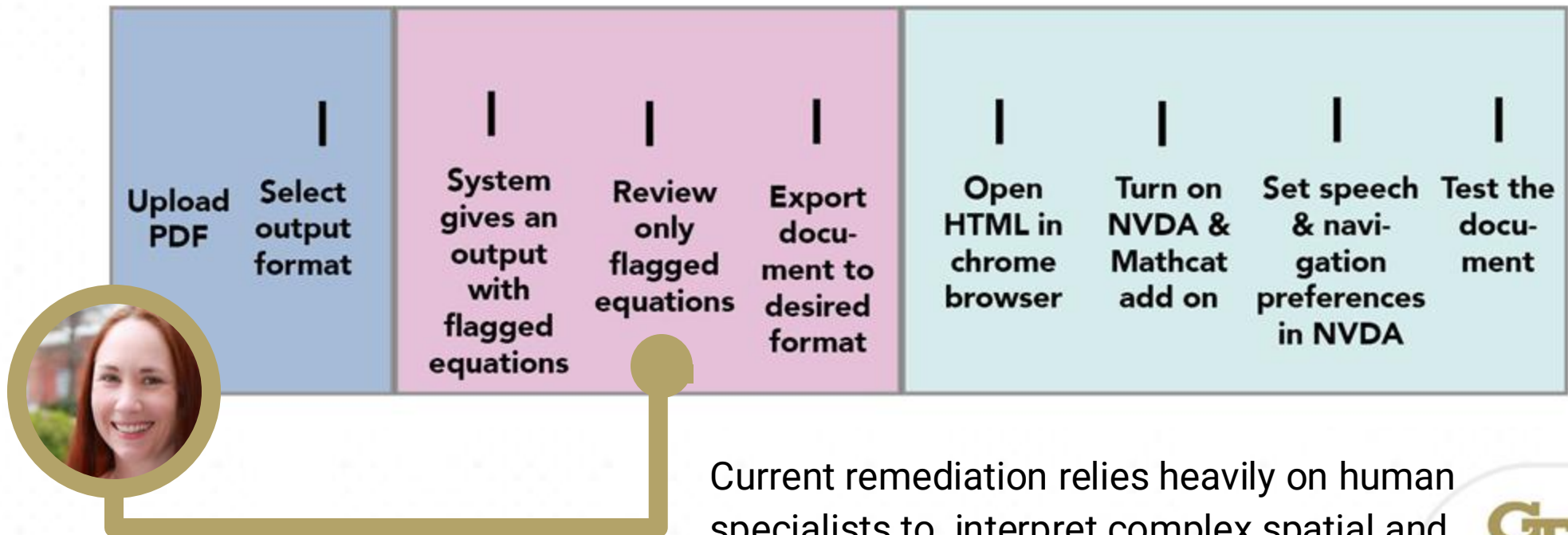
Current remediation relies heavily on human specialists to interpret complex spatial and symbolic information.

Problem Formulation

Dawn earned her bachelor's degree in psychology and her master's degree in social work from the University of Georgia.

Leads the E-Text Team in creating accessible documents for students with print-related disabilities, and has helped CIDI in achieving accessibility goals for the past 16 years.

Works closely with publishers to make their alternate formats available for students to succeed in their education.

Georgia Tech

Problem Formulation

Participatory Research
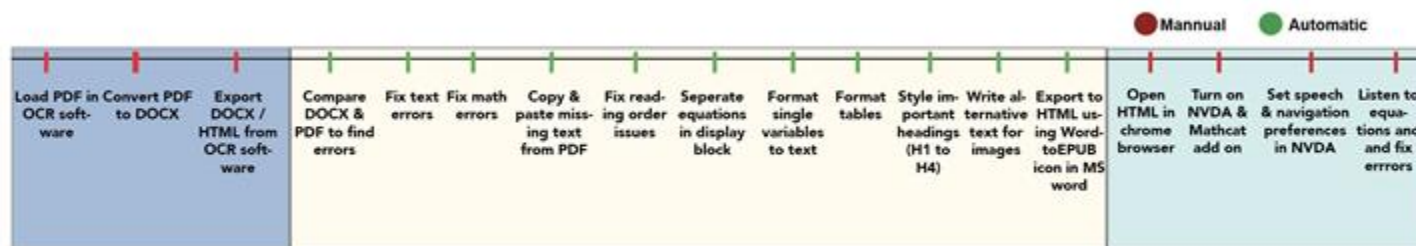
Tools used & Workflow Data

Identifying Current Limitations & User Pain Points

Chemistry Specific Knowledge

Challenges & Gaps in Chemistry-related Remediation

Alt-text generation Ratings

How will remediators add alt text for chemistry images without knowing SMILES?

● Mannual ● Automatic

| Load PDF in OCR software | Convert PDF to DOCX | Export DOCX / HTML from OCR software | Compare DOCX & PDF to find errors | Fix text errors | Fix math errors | Copy & paste missing text from PDF | Fix reading order issues | Seperate equations in display block | Format single variables to text | Format tables | Style important headings (H1 to H4) | Write alternative text for images | Export to HTML using Word-toEPUB icon in MS word | Open HTML in chrome browser | Turn on NVDA & Mathcat add on | Set speech & navigation preferences in NVDA | Listen to equations and and fix errrors |

Georgia Tech

Data for Training

Chemistry was the most novel area of STEM remediation research, but no available datasets on remediation content exist.
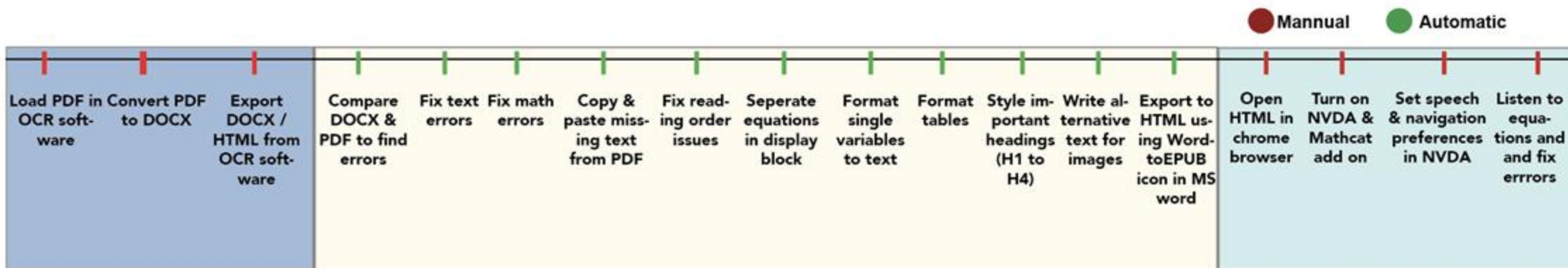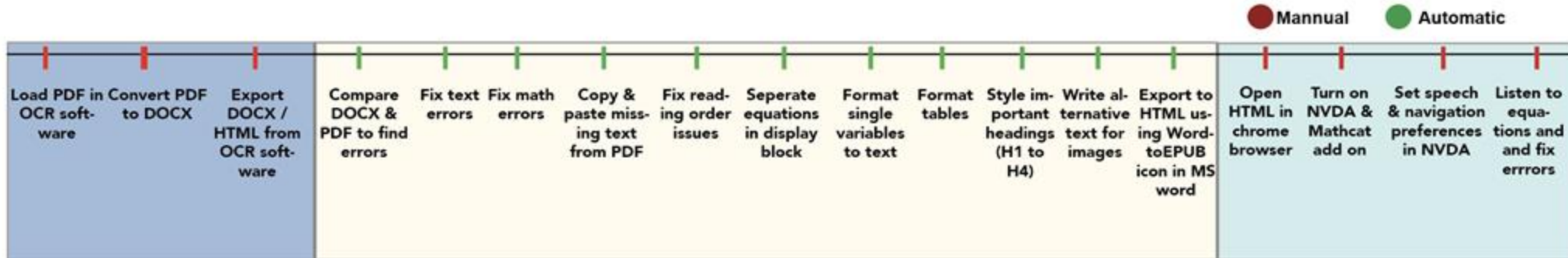
Dataset Challenge

Remediation in Practice and AI Ethics

Remediation Decision Making | Model Evaluation

Human-In-The-Loop Dashboards

Georgia Tech

Remediation in Practice and AI Ethics

Mannual   Automatic

Load PDF in OCR software | Convert PDF to DOCX | Export DOCX / HTML from OCR software | Compare DOCX & PDF to find errors | Fix text errors | Fix math errors | Copy & paste missing text from PDF | Fix reading order issues | Seperate equations in display block | Format single variables to text | Format tables | Style important headings (H1 to H4) | Write alternative text for images | Export to HTML using Word-toEPUB icon in MS word | Open HTML in chrome browser | Turn on NVDA & Mathcat add on | Set speech & navigation preferences in NVDA | Listen to equations and and fix errrors

IRB Protocol

Screener Survey, N=100+

Interviews and Cognitive Walkthroughs, N=30

Log data from real time evaluation, N=30

Human-Centered Model Evaluation and Decision Making

Live decisions with tools to use, prompts to use and models and refinements.

Model Evaluation Survey Questions

Human-Centered Model Evaluation and Decision Making

ChemGenAI

Start Here
    Consent
    Help
    FAQ

Task List
    Interview
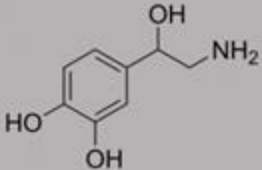   Evaluation
    Image 1
    Image 2
    Image 3
    Selection
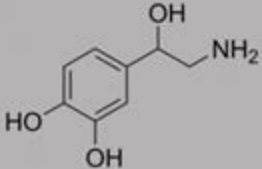
AI Description Evaluation   3A- Step 3

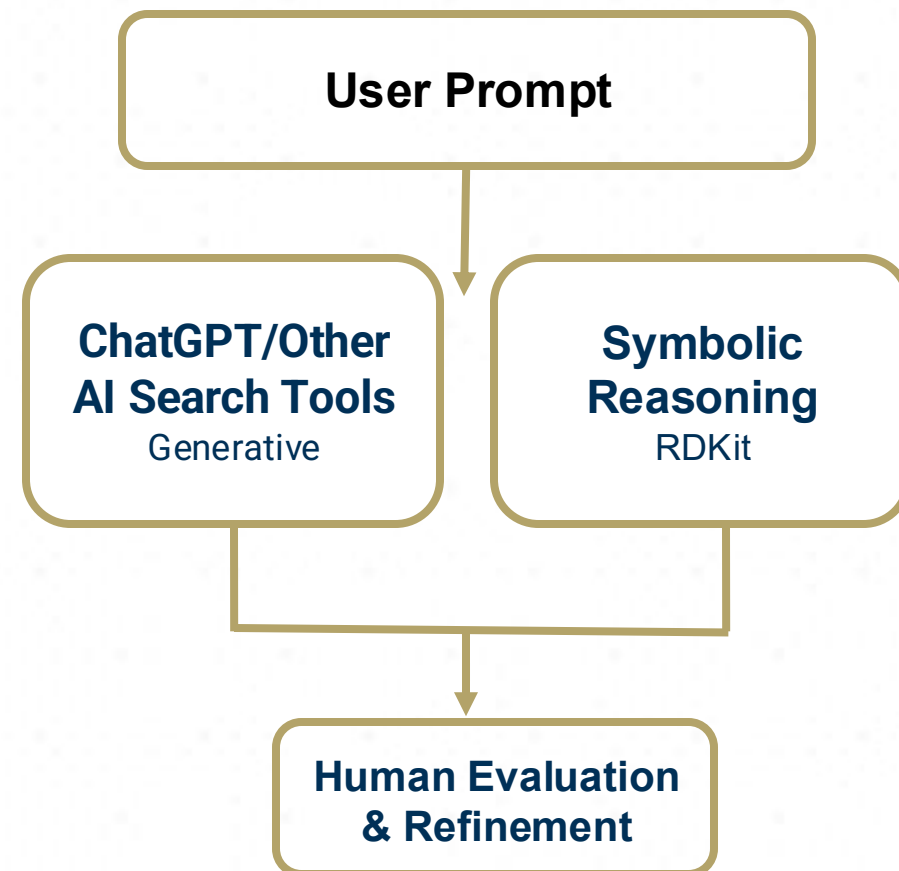Now please rate the AI-generated alt text on the following criteria (1 = Poor, 5 = Excellent):

Accuracy – Is the generated alt text factually correct?

Completeness – Does it capture all necessary parts of the image?

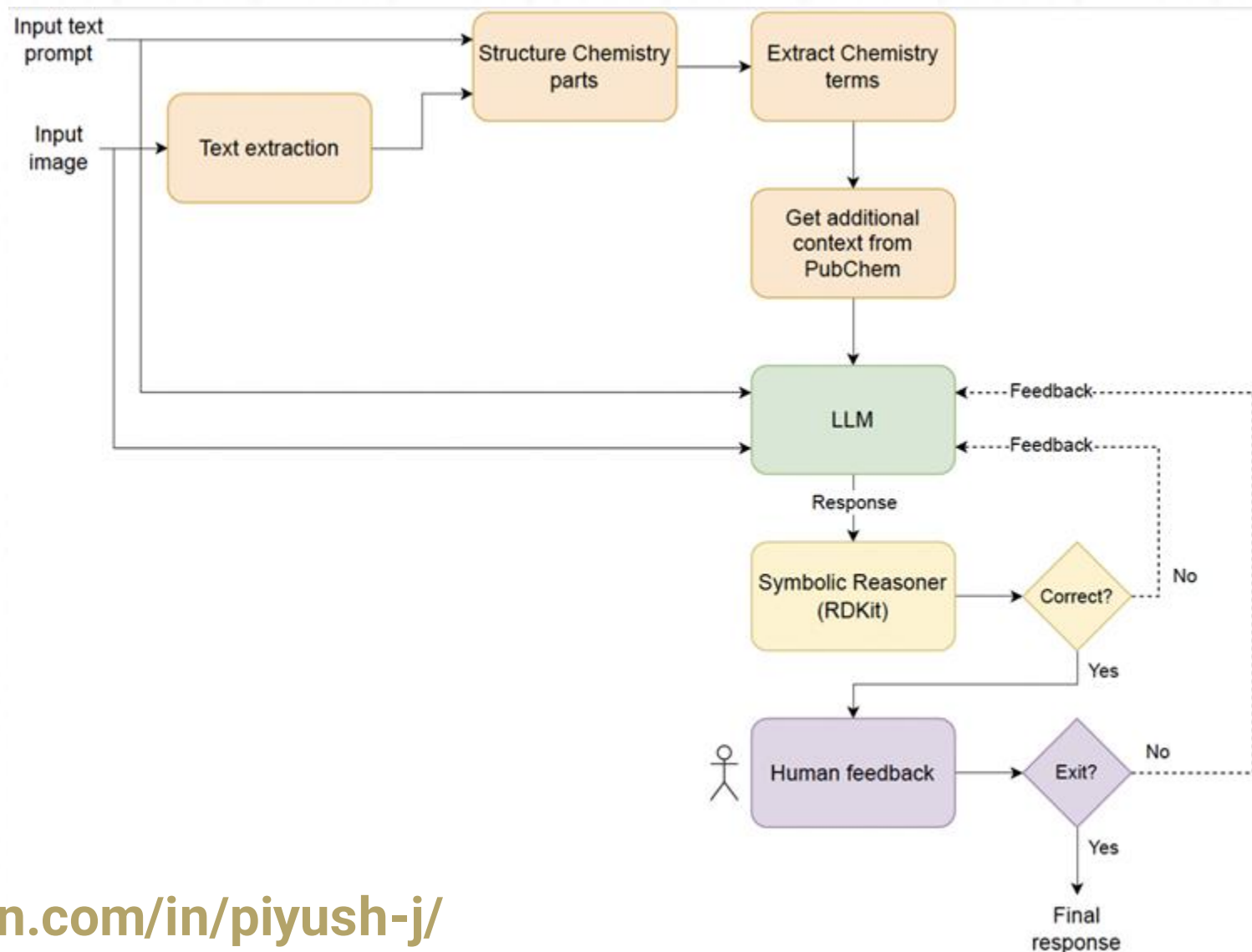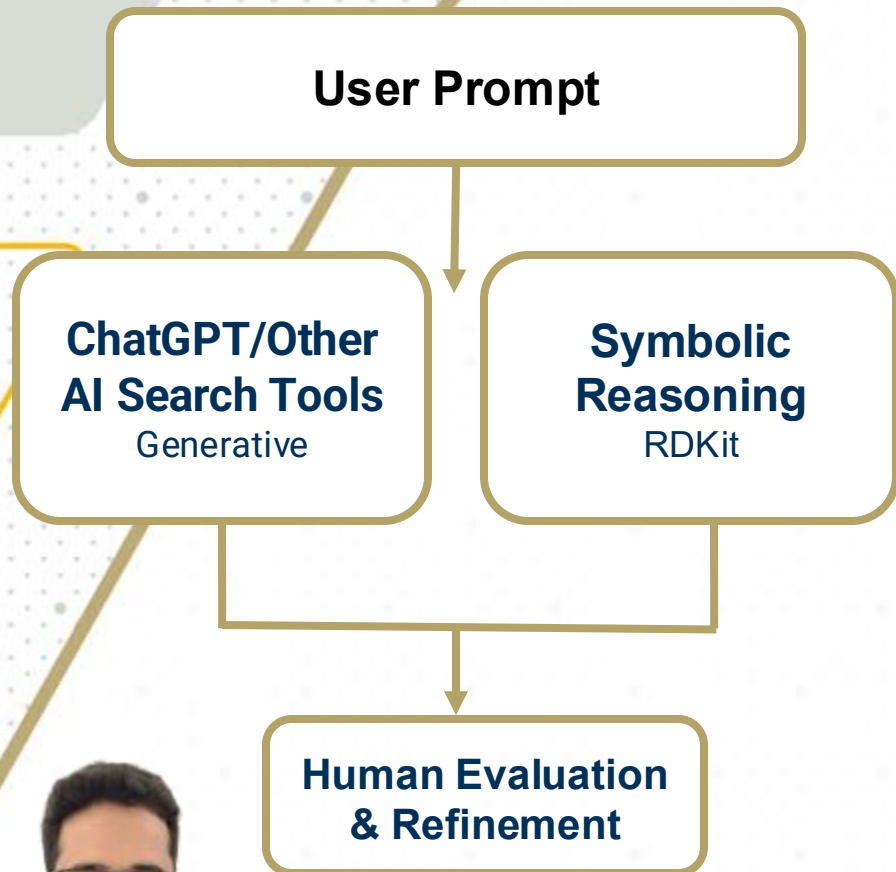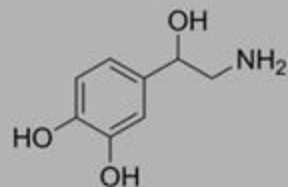Clarity – Is the language easy to understand and well-structured for BLV individuals?

Accessibility Best Practices – Does it follow alt text guidelines (concise, relevant, avoids redundancy)?

AI Generated Answer here

Model Evaluation Survey Questions

User Prompt

ChatGPT/Other AI Search Tools
Generative

Symbolic Reasoning
RDKit

Human Evaluation & Refinement

Live decisions with tools to use, prompts to use and models and refinements.

Georgia Tech

Remediation Decision Making | Model Evaluation

User Prompt

ChatGPT/Other AI Search Tools
Generative

Symbolic Reasoning
RDKit

Human Evaluation & Refinement

Human-In-The-Loop Dashboards
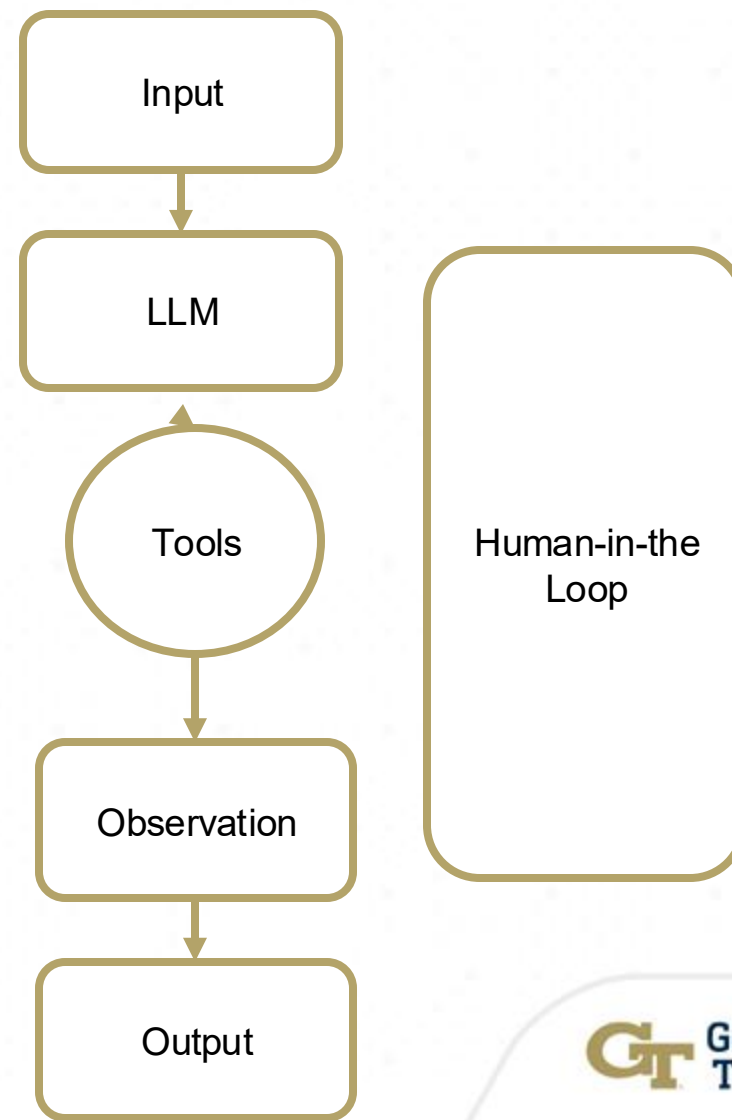

Model Evaluation Survey Questions

Next Steps

- Identify which equation types can rely on AI, which need partial human review, and which require full human remediation

- Use generative AI tools to produce descriptive alt-text for equations in each category

- Create measures to assess equation alt-text for both accuracy and readability, reliability across human-to-AI systems

Human-in-the-loop Feedback

Some errors need expert judgment, not just rules.

Real-world data is messy, and accurate correction is key for effective model training.

Symbolic Reasoning

Uses libraries like RDKit to verify chemical rules

Flags or fixes invalid SMILES

Offers token-level feedback for precise error correction

Training Generative AI to support chemistry learning and with remediation specialists.

Remediation in Practice and AI Ethics

Remediation Decision Making | Model Evaluation

Human-In-The-Loop Dashboards

Sumana Chilakamarri, Navya Ammayath, Guy Toles, Dawn Evans, Anand Nagpurkar, Maureen Linden, Piyush Jha, Vijay Ganesh, Abigale Stangl.

Georgia Tech