# Research Proposal

CADS
CADSLAB
cads@cadslab.site

Jingming Li
Nanyang Normal University
jmli@nynu.edu.cn

December 8, 2025

## 1 Project Title

**Using Thermal Comfort to Assess Cognitive and Perceptual Abilities of AI**

## 2 Background and Motivation

Thermal comfort reflects the human body's subjective perception of temperature, humidity, air velocity, and clothing insulation—a direct manifestation of the perception-decision loop under complex environmental information. Leveraging thermal comfort as an entry point allows simultaneous examination of an AI system's ability to sense multi-modal environmental variables, perform causal-association reasoning, and make adaptive decisions—areas that have not yet been systematically explored in existing AI cognitive assessments.

## 3 Core Hypothesis

> **If an AI can accurately perceive environmental variables, such as temperature, humidity, air velocity and provide reasonable thermal-sensation feedback based on clothing attributes, then its level of cross-modal reasoning, causal association, and adaptive decision-making can be quantified.**

## 4 Research Objectives

1. Build an experimental dataset based on the **ASHRAE Global Thermal Comfort Database II** for training and evaluating AI thermal-perception models.
2. Design a task chain—**environment perception** → **outfit selection** → **thermal-sensation feedback** → **adaptive action**—to test AI's cross-modal integration and causal reasoning capabilities.

3. Compare the results of this framework with existing **AI cognitive benchmarks** (e.g., MMLU, BIG-bench, GPQA, Humanity's Last Exam) to locate its position within the overall cognitive ability spectrum.

4. Propose a **thermal-comfort-based cognitive assessment metric system** to serve as a reference for future multimodal AI evaluations.

# 5 Literature Review

## 5.1 ASHRAE Thermal Comfort Database II

- Contains $\approx$ 81,846 fully documented field-experiment records covering 52 studies and 160 buildings, providing temperature, humidity, air velocity, subjective thermal votes, calculated comfort indices, and more.

- Uniform data structure and open access make it suitable for large-scale, standardized machine-learning training and validation.

## 5.2 Current AI Cognitive Benchmarks

- Recent high-difficulty benchmarks such as **GPQA, MMMU, BIG-bench, Humanity's Last Exam** have become key references for measuring AI "cognitive depth".

- These benchmarks focus on text, image, or code tasks and lack systematic evaluation of **environment perception-behavior adaptation**, leaving a research gap.

# 6 Methodology

## 6.1 Data Preparation

- Extract indoor environment records (temperature, relative humidity, air velocity) and corresponding **subjective thermal votes** from the ASHRAE DB II.

- Combine with **clothing attributes** (clothing thermal resistance, number of layers, material) to construct a three-tuple dataset: **environment – clothing – sensation**.

## 6.2 Task Design

Each step is compared against the corresponding real-world record in the database to compute **prediction error** and **decision consistency**.

## 6.3 Model Implementation

- Employ a **multimodal Transformer** (e.g., CLIP-style visual encoder + structured numeric encoder) to jointly process environmental numbers and clothing attributes.

| Step | Task Description | Target Cognitive Ability |
|---|---|---|
| 1 | **Environment Perception**: Provide AI with outdoor temperature, humidity, and wind speed values. | Multimodal sensing |
| 2 | **Outfit Recommendation**: Ask AI to suggest appropriate clothing combinations (layers, material) based on the given environment. | Cross-modal reasoning, causal association |
| 3 | **Thermal-Sensation Feedback**: AI reports its subjective thermal feeling (e.g., "slightly warm", "comfortable") using the selected outfit and indoor conditions. This step evaluates the AI's ability to generate realistic subjective thermal sensations. | Subjective evaluation generation |
| 4 | **Adaptive Action**: Based on the feedback, AI decides whether to **add**, **remove**, or **keep** the current clothing. | Adaptive decision-making, closed-loop control |

Table 1: Task Chain for AI Cognitive Assessment

- Integrate a **causal graph network** to capture relationships such as "temperature $\uparrow\rightarrow$ need for *fewer* clothing".

## 6.4 Evaluation Metrics

| Metric | Description |
|---|---|
| Environment Perception Accuracy | Mean Absolute Error (MAE) of predicted environmental variables |
| Outfit Recommendation Match Rate | Overlap with the optimal outfit recorded for the same conditions |
| Thermal-Sensation Prediction Error | F1-score for categorical thermal sensation classification |
| Adaptive Decision Success Rate | Proportion of AI decisions that lead to actual improvement in thermal comfort |

Table 2: Evaluation Metrics for Thermal-Comfort-Based Cognitive Assessment

## 6.5 Benchmark Comparison

- Run the same AI models (e.g., GPT-4, Claude-3, Gemini-1.5) on the thermal-comfort task chain and record the above metrics.

- Map the results onto the score ranges of **MMLU, BIG-bench, GPQA**, analyzing where the thermal-comfort task sits within the broader cognitive spectrum.

# 7 Expected Outcomes

1. **Open-access thermal-comfort-cognitive dataset** (environment, clothing, sensation triples) for community replication and extension.
2. **Thermal-comfort-based AI cognitive assessment framework**, including task definitions, metrics, and benchmark-mapping methodology.
3. **Peer-reviewed publications** (target journals: *Building and Environment*, *Artificial Intelligence*) reporting experimental results and analysis.
4. **Open-source codebase** implementing the multimodal perception-decision model, facilitating downstream research.

# 8 Project Timeline (2 months)

| Phase | Duration | Main Activities |
|---|---|---|
| 1. Data Preparation | 1 Week | Download, clean ASHRAE DB II; construct three-tuple dataset |
| 2. Task Prototyping | 1 Week | Design task chain; implement baseline models |
| 3. Model Development | 1 Week | Build multimodal Transformer + causal graph network |
| 4. Experiments & Evaluation | 2 Weeks | Run experiments; collect metrics; compare with existing benchmarks |
| 5. Result Analysis | 2 Weeks | Statistical analysis; draft manuscript |
| 6. Public Release | 1 Week | Submit papers; release dataset & code; produce final report |

Table 3: Project Timeline

# 9 Ethical and Societal Impact

- The study exclusively utilizes **public, anonymized data from the ASHRAE Thermal Comfort Database II**, avoiding any personal privacy concerns.
- Enhancing AI's ability to perceive and adapt to environmental conditions can benefit **smart buildings, wearable devices, and indoor climate control systems**, promoting energy efficiency and occupant comfort.

| Risk | Potential Impact | Mitigation |
|---|---|---|
| Data bias, like under-representation of certain climates | Skewed model performance | Data augmentation, stratified sampling across regions |
| Model over-fitting to specific buildings or populations | Poor generalization | Regularization, cross-validation |
| Inconsistent evaluation standards across benchmarks | Difficulty in comparative analysis | Adopt a unified normalization scoring scheme |
| Ethical compliance (use of human subjective data) | Legal/ethical issues | Use only publicly available, anonymized ASHRAE data; adhere to data-use agreements |

Table 4: Project Timeline

# 10 Conclusion

By using thermal comfort as a bridge, this project integrates traditional environmental perception with cutting-edge **AI cognitive assessment**, filling the gap in evaluating cross-modal causal reasoning and adaptive decision-making. Through systematic experimentation, rigorous benchmarking, and open dissemination, the work aims to provide a novel, quantifiable perspective on AI's general cognitive abilities.