Experiential Views

Towards Human Experience Evaluation of Designed Spaces using Vision-Language Models

Bon Adriel Aseniero
bon.aseniero@autodesk.com

Michael Lee
michael.lee@autodesk.com

Yi Wang
yi.wang@autodesk.com

Qian Zhou
qian.zhou@autodesk.com

Nastaran Shahmansouri
nastaran.shahmansouri@autodesk.com

Rhys Goldstein
rhys.goldstein@autodesk.com

Experiential Views is a proof-of-concept tool through which we explored a method of helping architects and designers predict how people might experience their designed spaces using AI technology based on Vision-Language Models (VLM). It evaluates a space using a pre-trained model (OpenAI's CLIP model) that we fine-tuned with photos and renders of a building.
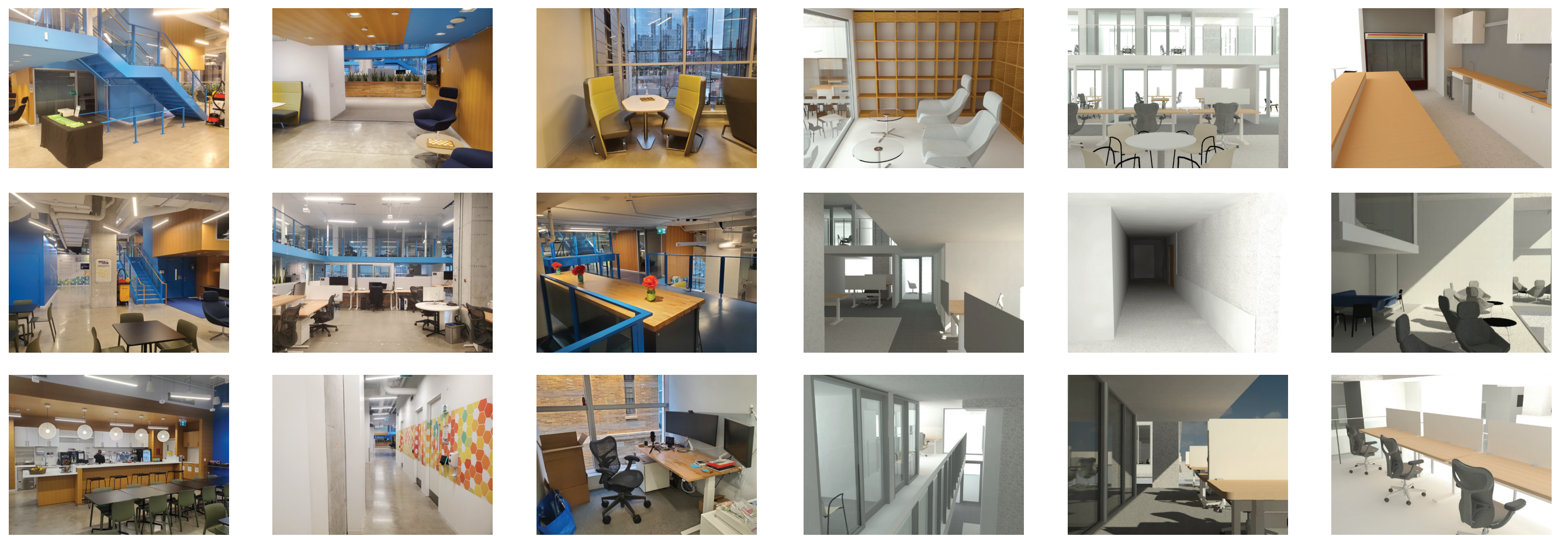
Example *"scenes"*—images and renders of a space used to fine-tune a VLM for Experiential Views.

Human Centric Dimensions

Experiential Views evaluates a space based on a preliminary set of three **human-centric dimensions** that characterize the Social, Tranquil, and Inspirational qualities of a scene.

The **Social dimension** comes from studies associating positive social interactions, high-quality gathering places, and a sense of community with positive outcomes for physical and mental well-being.

The **Tranquil** and **Inspirational dimensions** are loosely based on ideas from the biophilic design community, which encourages the creation of spaces that are *"calming," "relaxing"* and tend to *"reduce stress,"* as well as spaces that feel *"fresh, interesting, stimulating and energizing."*

Model Fine-Tuning

To fine-tune OpenAI's CLIP model for evaluating spaces based on our human-centric dimensions, we represented scenes as pairs of image and text; we conveyed the view of the space through an image (photo and 3D render), while we expressed the experience of the space as text (using our dimensions as keywords). We then let the model evaluate the scenes by computing the similarity scores between the images and text descriptions. We manually evaluated and classified 150 scenes: 100 of which were used for evaluation, while 50 were used for training.
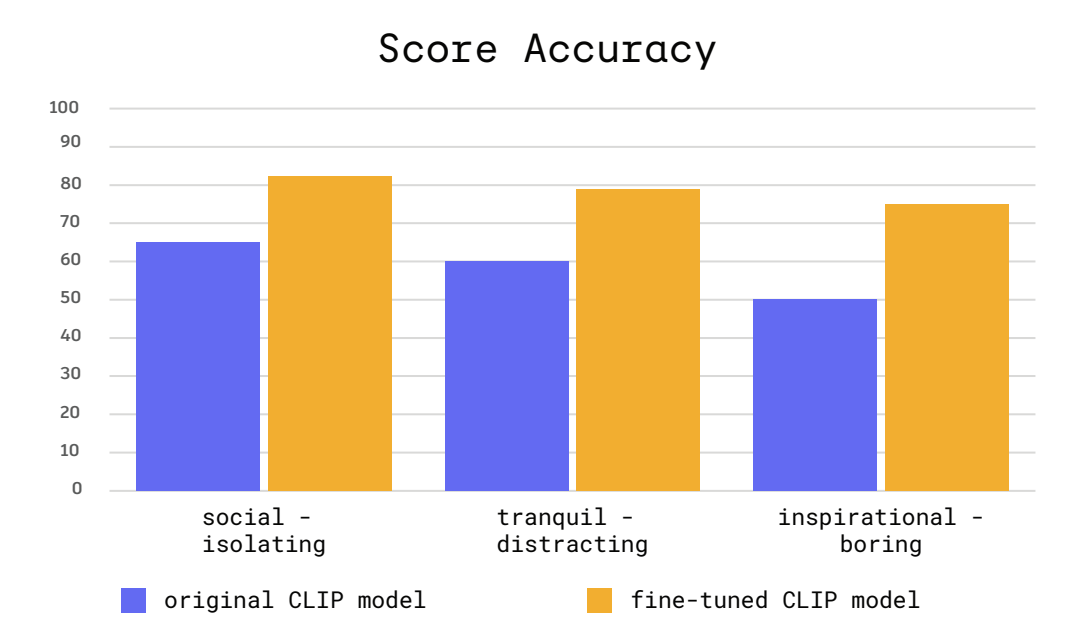
Accuracy scores computed using the evaluation set of 100 photos disjoint from the training set of 50 photos.

The accuracy of the scores relative to the manual evaluations was ~58% on our testing sets. After fine-tuning with 50 photos from our training dataset, the accuracy increased to ~78%.

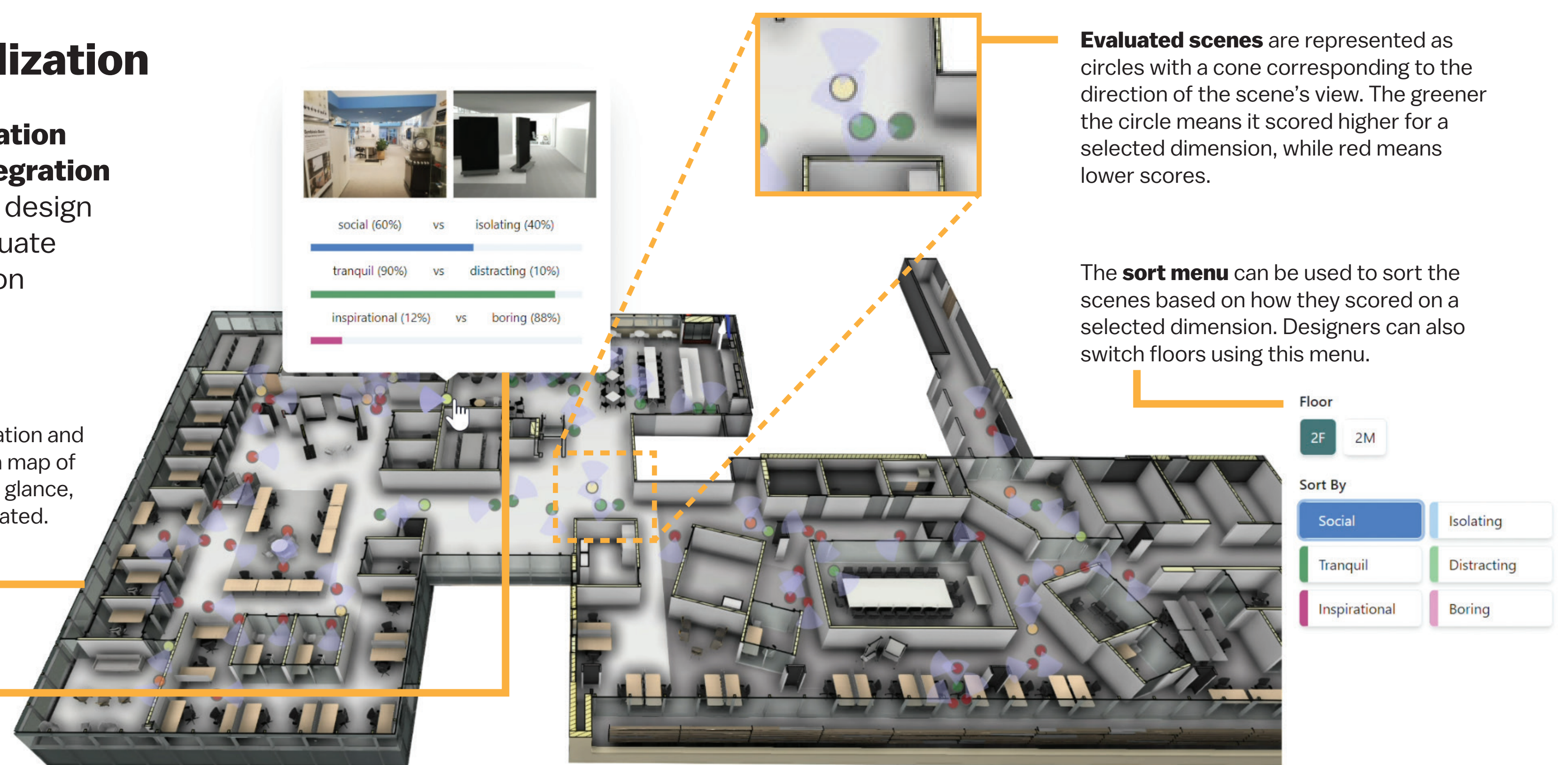## User Interface & Visualization

We developed a **floor plan visualization** and a WebGL-based **3D-viewer integration** that demonstrate how architectural design software could be enhanced to evaluate areas of a built environment based on psychological or emotional criteria.

Evaluated scenes are represented as circles with a cone corresponding to the direction of the scene's view. The greener the circle means it scored higher for a selected dimension, while red means lower scores.

The **sort menu** can be used to sort the scenes based on how they scored on a selected dimension. Designers can also switch floors using this menu.

The **floor plan visualization** displays the location and view cones of all of the evaluated scenes on a map of the space. This enables designers to see, at a glance, how much of the entire space has been evaluated.

More details about the scene are shown through a **tooltip** that contains the photo and 3D render of the scene and the scores produced by our fine-tuned model for each dimension (represented as bars).
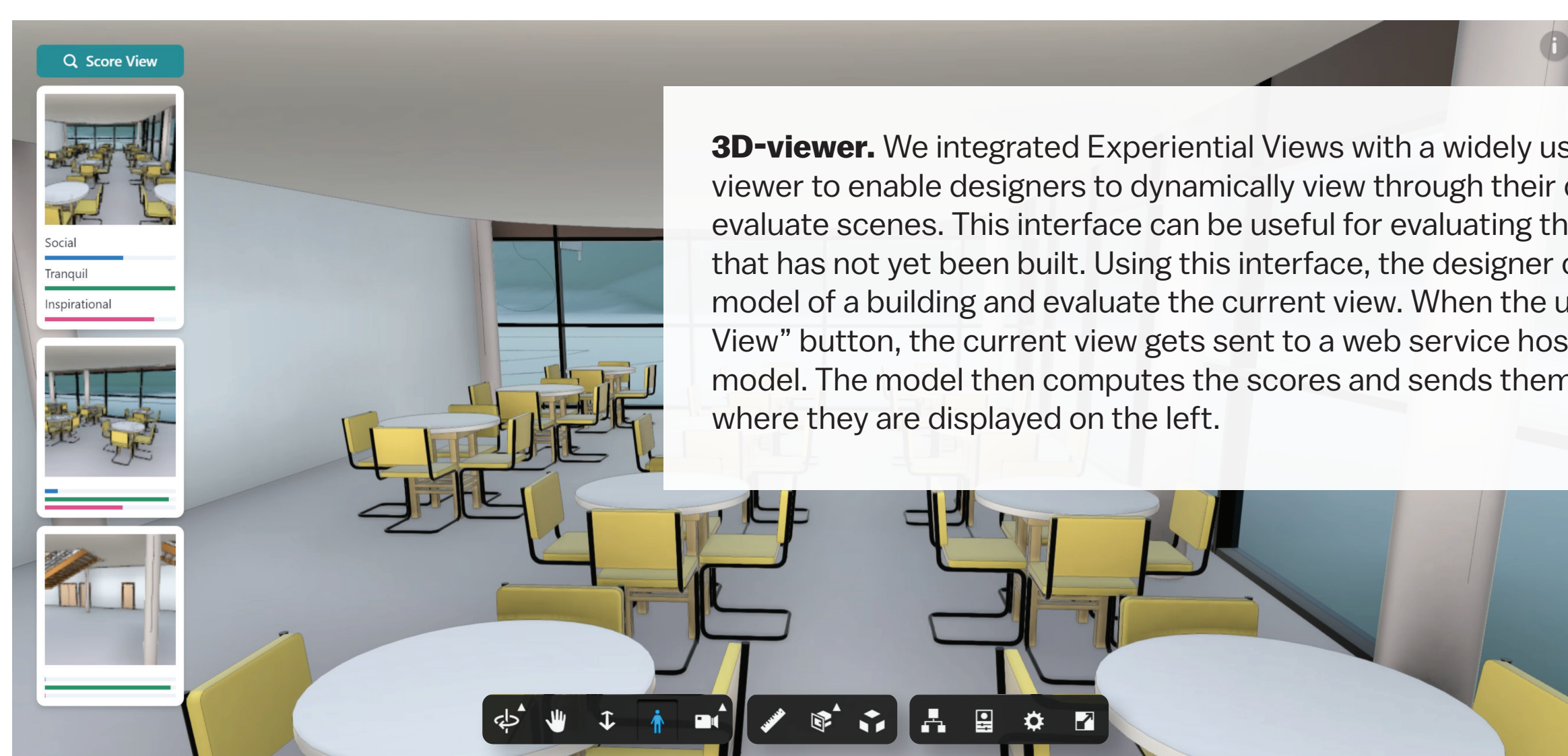
The **list of evaluated scenes** as icons containing the scene photo and score bars. Clicking or hovering over these icons shows their location on the floor plan by activating their corresponding circle and tooltip. Designers can sort this list by selecting an option from the sort menu.

3D-viewer. We integrated Experiential Views with a widely used WebGL 3D model viewer to enable designers to dynamically view through their designed spaces and evaluate scenes. This interface can be useful for evaluating the design of a space that has not yet been built. Using this interface, the designer can navigate through a model of a building and evaluate the current view. When the user clicks the "Score View" button, the current view gets sent to a web service hosting our fine-tuned model. The model then computes the scores and sends them back to the viewer where they are displayed on the left.

We see Experiential Views as an early step towards helping designers anticipate possible emotional responses to their designs to create better experiences for occupants.

AUTODESK
Research

read more about the project at
https://www.research.autodesk.com/publications/experiential-views/