Qualitative Pre-setup

1. Lay out on the desk by the tutorializing machine (in reverse presentation order):
    a. Participants receipts
    b. Goodness Questionnaires x3
    c. Explanation “Menu” to draw on
    d. Expertise Questions (#4-6)
    e. Ranking sheet add-on
    f. A stack of AARFAI forms (~15)
    g. Ranking Sheet
    h. DEMOGRAPHICS
    i. Facet Survey
    j. Expertise Questions (#1-3)
    k. This tutorial script!
2. On the presentation machine, do:
    a. Open the tutorial slides
3. On BOTH the presentation machine AND participants’ machine:
    a. Open Camtasia Recorder
    b. Open Pycharm, then in the study software repo:
        i. Click the blue arrow in the top right corner to pull latest version
        ii. Click the green play button to run
        iii. If it looks wrong, go to ThesisCode/src/settings.py
            1. On line 4, make sure self.enableDevControls = **False**
            2. On line 5, make sure self.enableHumanPlay = **True**
            3. On line 6, make sure self.enableNameObfuscation = **True**
4. Ensure that the desk has the following physical things:
    a. The cheat sheet (1 copy, to the RIGHT of the keyboard)
    b. Blank paper (10 sheets, to the LEFT of the keyboard)
    c. Pens/pencils, various kinds
    d. Finglonger, to point at stuff on their screen
5. Ensure that YOU have the following physical things:
    a. Payment
    b. Participant contact info, just in case
6. Perform a damn mic check, or risk being sad forever

Before Seating

Hi, may I have your name?

(if name is present, check their name off the list)

The study will run for approximately 2 hours, so if you want to use the restroom, now would be a good time.

Please finish any food/drink before entering the lab. You will be at this workstation with all the paper. As per CDC and Oregon Health Authority guidance, we require that you remain masked in the lab. The lab will **only** contain you and fully vaccinated and masked staff members, numbering no more than 1. Thanks for participating today!

If the participant needs a break

(Let them, *note the time somehow so we can correct the timing records, then point the way)*

Introduction

Hi, my name is <<insert name>> and I'll be running today's study. First, please turn your phone **all** the way **off**, like this!

(demonstrate and wait for participant to DO it)

I'll be reading from this script to be consistent with you and all the other participants.
Please **don't discuss the contents of this study with anyone**, as we don't want **other** participants to receive any information in advance.

This study is about looking into how Artificial Intelligence (which I'll abbreviate **"AI"**) can be explained.
We want to make AI more understandable for humans like you and I; your participation will help us better understand how to do this, so **thanks** in advance!

Your task today will be to **assess** different AI agents (or just "agents") playing a game, where each agent has their own strengths and weaknesses.
These agents aren't perfect---sometimes they'll provide inaccurate estimates and make bad choices.
Your objective will be to rank these agents from the one you think is the "best" agent to the one that's the "worst".
Think of the ranking as like a draft board in your favorite sport or as a list of candidates you might like to hire.
Each one might have their pros and cons, and your first choice might not always be available.
As a result, it's important to be as accurate as possible throughout the list.

Next we'll explain the **game**, the **agents**, and the **explanations**.
Before doing so, please click the red "record" button to start the recording.
This will capture onscreen activity and words that we say only, there is no video.

Learning the Game the AI Agents Play

So the AI agents will play a game **like** Tic-Tac-Toe, only bigger.
If you look at my screen, you will see an example of a game of Tic-Tac-Toe.
In this kind of game, two players take turns placing their mark (X or O), and try to get 3 in a row on a 3x3 board.
And so here we see X has gotten 3-in-a-row vertically, and so wins the game

(show slide with Tic-Tac-Toe, play the animation, then advance to next slide, with 9-4-4)

In the bigger game our agents play, two agents will try to get **FOUR** in a row on a 9x4 board.
As in Tic-Tac-Toe, a winning sequence of **FOUR** in a row can be horizontal, vertical, or diagonal (as we see here).
So let's take a look at your screen… in this part of the window, you'll see the **games** play out.

(*Point to the game area)*

I'll have you play this game to get a feel for it.
It's currently blank, so let's create a game---but first, we need some **agents** to play the game!

Each agent will have a **color**, which serves as their **name**.
Up here, you'll see a list of 6 **tabs** for the Blue, Green, Orange and other agents.

(*Point to the game area)*

When you select one of these **tabs**, that is the agent that you choose to **assess**.
This means two things: **first**, that agent will be playing **X** for its games, and **second**, that agent will show you **explanations** for its decisions.
You will **not** see explanations for decisions made by the **opponent**, which will **always** be playing O.
Upon reaching the main task we will replace these agent tabs with new ones.
Go ahead and click on the Green agent tab...
As you do it, notice how in the top-left, we now have the Green agent **listed as playing X**.

(point to the X player static text)

The Green agent needs an opponent to play O, which will be **you**.
To accomplish this, click on this **drop-down menu**… select the "**human**" option.

(Point at the O player drop-down menu)

Once you have game settings we like, you can click the **step** button here to create the game, and you should see an **empty board** appear.

(Point at the step button)

So now we have made a **game**, with two agents and an empty board.
When the AI makes a decision, you'll see charts appear at the bottom, but please ignore them for now; we'll go over how to read them later.
When we click "**step**", the AI will make its decision, so go ahead and click step again.

(ensure they click step and see a a piece appear)

Now, which of the squares would you like to take?
Go ahead and click on it... you should see it highlight in **black**.

(ensure they indicate a desired move)

To **submit** that move, you'll have to click "Step".
Go ahead and do so when you have selected a move you like.

(ensure they have MADE a move)

Now you have **everything** you need to play against this agent, so... go ahead and finish the game!

(wait until they finish the game)

First, I want you to notice that when a player **wins**, the 4-in-a-row will **light up**.
Second, I'm going to have you answer a few questions here.

(Give them pre-expertise questions 1-3)

For these three sheets, choose from A or B which move you think is better for the situation depicted.
Go ahead and complete all 3 by writing a letter or circling a move.

(Wait, then collect expertise questions 1-3)

Third, please answer these questions about your interaction with AI-powered technologies.

(Give them Facet Survey)

What we mean by AI-powered technologies are things like Amazon's Alexa, PowerPoint's Design Helper, Apple News, and even this game you just played.

(Wait, then collect expertise questions 1-3)

Creating Games between Agents

Your task today **isn't** going to be playing against different agents.
That was just a fun way for you to familiarize yourself with the game.
Your task is to **assess** different agents and to rank them from **best** to **worst**.
So we already have the Green agent, which you played against, but it needs an AI opponent to play the O shape.
Click on that drop-down and leave the menu **open**.

(Point at the O player drop down and wait for them to open it)

Then we can see that **other** agents are listed there, along with the "human" option which we just used.
Go ahead and select the "Orange" agent… and notice that **indeed**, it is now the Orange agent beside the O.

In games like tic-tac-toe, it's **advantageous** to go **first**, which means the player who goes **second** is at a **disadvantage**.
That's why here… We've provided you a **toggle** to change who goes **first**

(Point at the player toggle)

I think we want to see how well the Green agent does if it goes second.
Go ahead and click on the toggle, and notice the change... it says that "O player" will play first.
So, that's how you'll be able to **tell** who's going first or second, as well as **adjust** that setting.

In order to see the game's progress, you'll be clicking on the "**step**" button there to see a single decision.
Go ahead and click on that button **two or three** times, stopping when you see an **Orange** piece appear.

(verify they are on the rails)

Now look at the Orange agent's first piece. in the top-right of that square with the O, you'll notice a "1".
That tells you the **decision number** of the game.
At the bottom, you'll also see that a chart appeared.
We'll cover this more in-depth later, but that's going to become one of the explanations that the agent generates to show its reasoning.
It doesn't have any data on it because the agent you're assessing hasn't made a **decision** yet.
Once the X player makes a decision, you'll see data appear on the explanation.
To be clear, that means you will **ONLY** see explanations for decisions made by the agent that is playing **X**, in this case the Green agent.

Alright, can you advance the game forward with the “Step” button until around decision number **6**?

(verify they are on the rails)

Notice at the **bottom** of the screen, there’s a status bar that shows the current decision number.
Out of curiosity, which of the agents do you think will win at this point?

(wait for an answer)

Alright, please click the “Step” button until we find out!

(verify they are on the rails)

Now, I’d like to go over this **slider**, move it very slightly and you should see a piece or two disappear.

(point at the rewind slider)

The purpose of this slider is to let you “**rewind”** and see how everything on the screen changed as time went on.
It does this by **filtering** the information on the screen, showing only what information was available at that time.

To learn this feature, you may find it helpful to start a new game and play with the slider repeatedly while stepping.
Go ahead and drag the slider around a bit to get the hang of it, let me know if you have any questions.

(verify they are on the rails)

How the Agents Choose their Moves

(IF there are not some pieces on the board) I want you to click “step” a few times please to start a new game and get a few pieces on the board.

(verify they are on the rails)

Let’s discuss how the agents make decisions.
First, it looks at the current game **board**.
Then, for each possible **move** it could make, it calculates what it thinks are that move’s probability of eventually:

- **Winning the game;**
- **Losing the game;**
- and **Drawing the game.**

The agent then gives each move a **score**: the **Win** probability minus the **Loss** probability.
Thus, a high score means the agent thinks it is more likely to win if it makes that move than a move with a lower score.
Note that illegal moves should be assigned very low scores.
The agent **won’t always** select the action with max perceived value, but a selected action **will always** be **near** that max.

That said, we have prepared 3 explanations for you to inspect these scores.

*(Ensure **scores through time** is the shown explanation)*

Expl 1: Scores Through Time

The first explanation (Scores Through Time) uses **time** as the X-axis

(point at the X axis at the bottom)

...and every time that the agent you’re assessing makes a decision, a new **column** appears in the visualization, labelled with a number at the bottom.
So here, we are at decision number <number, should be 3> and about to see decision <number, should be 4>.
Click step for me please.

(point at a column and then the label)

Notice that in each column, there’s **one** rectangle the same color as the agent, in this case **GREEN,** but that there’s also many grey rectangles in each column.
The colored rectangle is associated with the move the agent **selected** for that decision point.
There are a total of 36 moves on the board, so grey rectangles are the scores for the 35 moves that did **NOT** get selected, even the illegal ones that aren’t empty.
Notice that the illegal ones have really low scores.

(point at some illegal scores)

However, scores that are higher than this are considered better, with a score up here being maximum, corresponding to a 100% chance to win the game.

(point at where high scores are/would be)

We also have tick marks on the Y-axis to help you understand where these scores are, with major ones at intervals of ½ and minor ones at 1/10.

(point at tick marks)

So now it's your turn… Place your mouse cursor over any **square** on the board and leave it there.

(verify they are on the rails)

Notice that it will highlight the appropriate **scores** for every decision through time. So, to be concrete, this score which is now highlighted in **yellow** corresponds to the score the agent gave to taking this square at decision 2...
Meanwhile this is the score for the same move later, at decision 4.

(point at a highlight, then another)

Now, place your mouse cursor over any **score**, the rectangles in the explanation**.** Notice how it highlights the moves on the board associated with that score. Because many moves can be given similar scores, often many scores will be drawn on top of each other, so the selection may **not** be unique.

(highlight a score which has overlapping)

Now, click step and observe how the explanation changes after the agent makes a decision, namely that a new **column** appears for that decision point.

(verify that they click step, and they might need to do it a second time. Point to new column.)

Please take some time to explore this explanation and ask any questions that arise.

(wait until they seem satisfied, or… ??? time)

There are 2 more explanations for these agents.
We have been exploring the scores through time so far, let's look at another one.

Expl 2: Small Multiples

To see the a new explanation, which we call Small Multiples, go ahead and click on this button

(point at the small multiples radio button)

The Small Multiples explanation behaves similarly to the previous one, **except** it divides up the same information spatially.
That is, for each square, there is a Scores Through Time chart.
To be concrete, this **top left chart** here shows the agent's perception of how likely it is to win if it placed an X in this **top left square** at any previous decision.

(point at the chart, then the square)

The same is true of **this** square and **this** chart, and so on

(point at another chart, then the corresponding square)

Each of the **columns** from left-to-right are the same as those in the Scores through Time explanation, which we represent with the same **coordinate axes** and same **decision number** labels (though they are tiny now).
So the far left of each chart represents the score X gave to that move at the **first** decision, and the far right represents the score X gave to the **same move** at the most recent **decision**.
If a move was selected by the agent showing explanations at a certain decision point, we will render that move in the agent's color.
So in this case we see a highlighted score here, because the agent **selected** that move at that decision point, which we can see on the board here

(point at a highlighted score, then the corresponding square)

Now it's your turn…Mouse over any **square** on the game board that you'd like.

(verify they are on the rails)

Notice that it highlighted the **chart** corresponding to that square.
Now, how about you place your mouse over the chart that's highlighted.

(verify they are on the rails)

Notice that it highlighted the same square we had just previously been hovering!
This highlighting will help you see the spatial alignment of the grid of pieces and grid of charts.
Now, click step and observe how the explanation changes, namely that each chart gets a new data **point** for the decision that just occurred.

(verify they are on the rails)

Please take some time to explore this explanation and ask any questions you may have.

Expl 3: Best-to-Worst

There's one last explanation I'd like you to see.
Look back at the explanation controls on the left, and you'll see an option for the "Best-to-Worst" explanation.
Go ahead and click on that.

(verify they are on the rails)

In this explanation, the Y axis is the same as in the first explanation, but the X-axis is no longer time, instead it represents the sorted ordering of move scores at a single decision.

(point at Y axis then X axis)

So in this explanation, each decision point generates a single data series.
When the series is first generated, we show it in the agent's **color**.
As the game progresses, old lines will become **grey** and increasingly **transparent** as they age further.
This means a dark grey data series is from a recent decision, while lighter ones are older.
Each data series contains the scores of every move on the board---even illegal ones---meaning each will contain 36 rectangles.

As with the other two explanations, you can mouse over a square on the game board, so go ahead and do that.

(verify they are on the rails)

And you can see that the score associated with that move is highlighted in every data series associated with previous decision points.

However, this explanation is a little different than the other two in some ways.
You can **only** interact with the rectangles that are the same **color** as the agent.
Hover over the leftmost of them down here in the explanation.

(Point at the best action)

Notice that when you do that, the square associated with that score is highlighted.
Within each series, the explanation orders the scores from left to right in best-to-worst order.
Currently, we are highlighting the move the agent felt to be best at the decision that **just occurred**.
Now highlight the rightmost <green> score.

(point at the worst action at this decision)

This score is the one associated with the move the agent felt was the worst---a certain loss.

Look at the highlighted square on the board. Is that a reasonable score for that move?

(wait for an answer)

Now highlight the square on the board where the **<Nminus1th>** piece landed. Observe the location of the highlighted square in the previous data series---it had **near-maximal perceived value**, and that is why the agent selected it!

(verify they are on the rails)

While the sorting used in this explanation can help reveal certain information, as a result of doing the sorting, if you want to find a specific move, you may need to hover some scores and/or squares.
Now, click step again… and notice that the explanation changed!

(verify they are on the rails)

Specifically, the **colored** series from the last decision became a dark **gray**, the **older** scores got a bit **lighter**, and a **new** colored series appeared.
Feel free to explore this explanation a bit more, and feel free to ask me any questions you might have at this point.

You will find a cheat sheet on the right of your workstation, which summarizes the explanations we just went through together in case you forget anything.
However, at any point in the study, feel free to ask any questions you may have about the interface.
<<Give some time to review the cheat sheet and ask questions>>

AAR/AI and Thinking Aloud

To remind you of your task, your job is to **assess** several AI agents and **rank** them.

To assist in your assessment, I will be giving you a form to fill out after each game.
Broadly speaking, the form asks you to consider a series of questions about what you observed, such as: **What** happened, **Why** did it happen, What **changes** can we make to do better?

We'd like you to "**think aloud**" while you assess these agents, because it will help us improve the system.
To do that, just say everything you're thinking and noticing as you go along.

Let's practice. Please think aloud while you figure out:
"How many windows are in your childhood home?"
It seems like you've got the idea.

Ok, here's another practice. Please think aloud while you figure out: "What are all the US states you can ski in?"

Main Task

With that, we are done with the tutorial.
Now, you're going to be ranking **SIX** agents.
We've given you this sheet to help you keep yourself organized.
(provide the ranking sheet)

You can use this sheet to write down preliminary rankings, jot down notes, whatever you need. There is more blank paper to the left of your workstation should you want it. Near the end of the study, I will ask that you write down a final ranking on this ranking sheet.

I'll chime in at various points with suggestions of things you may want to investigate next and provide forms at the appropriate times.

Remember to think aloud!

Whenever you are ready, go ahead and press the **"start"** button in the lower left corner to begin.

Main Task Guidance

(After first game) If you can, try to think aloud while answering;
It will help us understand your thought process. If it's too distracting to talk and write, feel free to explain your reasoning after you answer. I'll also mention at this point that the agent names have unique first letters, so if you would like to abbreviate agent names with a capital letter, that is fine.

(After every game) Ok so let's fill out one of these forms. In that game we saw A vs B, with ? going first, and ? winning. What did we see in that game?

(After finishing a form) Ok, done with that form? Would you like to rank/rerank or start a new game? If a new game, what configuration would you like?"

(Each time they rerank: note the game number after which the ranking occurred. For questions, alternate between...)
Your current ranking has <agent> at the **top**.
In what circumstances might you prefer a different agent??
OR
Your current ranking has <agent> at the **bottom**.
In what circumstances might you prefer another agent **less** than that one?
(if they ask to clarify, "you can define it however you like)

(If the thinking aloud lapses) Is there anything in particular you are looking for right now?
Do you find the scores you are seeing credible? E.g. do good moves have higher scores than bad moves?
Can you tell me what you were thinking about as you just did <whatever they just did>?

Post-Task Questionnaire

Great! We've completed the main study. Please click the "done" button to stop the timers.

First, can you tell me what your next steps would be to assess these agents if you had infinite time available for the task?

(Wait for their response)

Second, please provide the following ratings about your ranking, this is intended to be used as a long paper, like so...

(Provide the ranking sheet add-on, and collect when done)

Third, we have a few questionnaires for you to complete...

(Provide the expertise questions and collect when done)

Fourth, we have a few questions based on your experience in our study. As you answer, feel free to use any of the pencils/pens on the desk to draw on the explanation to show us any changes you might like. *Note: Can follow up any of the questions to come with, "Could you elaborate on that?" or "Is there anything you wanted to add to that?" or "Can you explain more about that?"*

(Place all three explanations in front of the participant)

(Time Permitting, have them do Goodness Questionnaires)

Can you point out which of these explanations was your favorite one?
… What about this one made it your **favorite**?

Alright, so which one of these remaining two was your least favorite?
… What about this one made it your **least** favorite?

So that leaves this one in the middle.
… What about this made it **better** than that one but **worse** than the other one?

Is there any way you would have liked to interact with any of the explanations that we didn't provide?

Any other comments or suggestions?

Before we get you outta here, would you like to see the right answer to the ranking of these agents?

If No, proceed to Post-study payment procedure. If yes:

Ok! Go ahead and minimize that window.

My screen will show the large scale tournament results and ranking that would result, based on the agents' "private" names.
Your screen now shows the mapping between private and public names, so the right answer is <answer>..

<summarize where the differences in their errors were> Again, please don't discuss this study with anyone, as we don't want other participants to receive any information in advance.

Post-Study Payment Procedure

Great! This completes the study. Please don't discuss the contents of the study with anyone, as we don't want other participants to receive any information in advance.

Please sign this receipt stating you received your payment for participating and I will trade the receipt for payment. Thank you for your time and assistance, have a nice day!

Post-Study teardown

1. Stop recording
2. Start processing videos
3. Close python software **with the X** (might need to wait a minute while it closes files. Closing the software any other way may cause log truncation)
4. Tag ALL documents with PIDs
5. Save raw/processed videos onto the server here:
    a. http://guille.eecs.oregonstate.edu/burnett/iftdata/2022-IUI-JonThesis
6. Copy usage log onto the server, named with PID