# Future of Desktop Human Computer Interface [HCI] 2025

In the future, AI will do knowledge work that is normally done by people today. Advancements in spoken/written language comprehension/generation by AI technologies will open a new interactive channel between computers and humans.

This article forecasts how desktop computer HCI will change. The key points are:

1. **Primary output** will continue being the **screen** because people read information much more quickly then they can listen to information.
2. **Secondary output** will be **computer generated speech**. Speech output can be used whenever users are visually occupied. It can deliver audio explanations of onscreen images, videos, or documents so that users can listen to explanations of visual items.
3. **Primary input** for future apps that complete large workstreams will be **speech** because speaking is faster than typing. Users will give computers a mission to accomplish rather than detailed step by step instructions. Users will ask critical questions about how the work was done so that users can build trust and confidence in the work. Revisions will involve changing large swaths of work with verbal guidance.
4. **Keyboard** and **mouse will remain** as input devices so that human users have the option to fall back on today's HCI as necessary.
5. **Spoken** questions/instructions will be more accurately understood by the computer if it knows the target onscreen object of user inquiries. Users can convey this contextual information by selecting items with **mice** and then speaking or alternatively, by letting **eye tracking devices** identify what users are looking at onscreen as they are speaking.

# Table of Contents

# Trends That Will Shape Desktop HCI

Today, a desktop computers' standard output is a screen, and the standard input is a keyboard.  Mice provide fine motor control to interact with GUIs.

As the nature of the work that AI completes for human users evolves and new computer input/output technologies grow more reliable, the standard human computer interface [HCI] will also change.

Trends that will change HCI:

1. Nature of AI Work: As AI …
    a. … get tasks **done independently**, human users will want to review the steps agents took to complete the work so that they can trust the work.
        i. "How did you acquire this information?"
        ii. "Which online stores did you check for the price of X?"
    b. … assemble **collections of options** such as shopping items across multiple websites, then human users will want to review the work in a single uniform GUI rather than visiting the original websites that have the options so that the collected information is easier to consume. Users can make the purchase on the original website.
    c. … explain **complex concepts**, users will want charts, illustrations, animations, and presentation slides to ease their comprehension.
    d. … **edit blocks of content** including text, illustrations, charts, photos, videos under user guidance, then users will want to view content as before/after views and sometimes see all revision possibilities/choices all together.  Human users will want the history of the changes and have the option to undo them.
    e. …get more **complex tasks done**, human users will want to ask critical thinking questions as supervisors ask their direct reports.  "Did you consider the alternative interpretation of X as ?" "Did you try looking for quantitative data that backs up the argument?"
2. Technologies
    a. Improvement of **speech driven input/output** channels will enable users to speak to and listen to AI.
    b. **Eye focal point tracking** that tracks what onscreen item that users are looking at will enable AI to more easily understand the context of users' spoken instructions and questions.

# Today's Desktop HCI

This is how human users use screens, keyboards, and mice when getting work done.

## Output - Screen

Users consume information on screen as text, photo, video, charts, illustrations, and animations on screens.

Users get visual feedback when the mouse pointer alters from an arrow to a hand/text cursor that indicates what areas are receptive to mouse clicks.  In addition, the onscreen object itself can alter its appearance when the mouse pointer is on top of it.

Users interpret icons, read text labels on buttons, on-hover labels/information to anticipate what actions clicking the object or button will trigger.

## Input - Mouse, Keyboard

Mouse pointers enable fine interactions with onscreen objects with clicking, hovering, holding, dragging:

1. **Selection** of objects, text, parts of photos, clips of videos.
2. **Precise cursor placement** into an area of text.
3. **Triggering action** with click.
4. **Scrolling** vertically/horizontally with two finger touch (available on Apple Magic Mouse).

Keyboards enable:

1. **Text** entry.
2. **Action short cuts** such as Command-X for "cut."
3. **Submission** of information with "enter" key.

# Tomorrow's Desktop HCI

Let's review how new speech input/output channels and eye tracking can affect HCI.

## Speech Input/Output, Today's Apps

A speech output channel does not accelerate the consumption of information because people can read information faster than they can listen to it.

A speech input channel can accelerate the input of text by allowing human users to enter by speaking instead of typing with no typos. However, it does not accelerate, for the most part, mouse interactions of selection, cursor placement, and triggering action. Verbally describing what to select or where to place the cursor will likely be more arduous than using a mouse. Moving a mouse and clicking, or pressing shortcut key combinations or the "enter" key is likely faster than speaking the equivalent commands.

The existing screen, mouse, and keyboard interactions will likely remain for today's applications because the speech channel alone does not help the user get more done efficiently. This result is not surprising because today's apps are built for the screen, keyboard, and mouse.

## Eye Tracking, Today's Apps

Eye tracking can substitute the mouse movements as long as precise control is technically and physically possible. A button that substitutes for the buttons on a mouse is necessary.

Human users don't reposition the mouse pointer unless some action such as cursor placement is necessary to switch applications, edit content, or trigger action. Sometimes, users use mice as a way to trace through lines of text as they read the screen.

An eye tracker that tracks all the human users' eye movements may be annoying to some users. To avoid this, human users can use a single finger touch on a button to activate the eye tracker to act as a pointer. Hover effects are triggered with single finger touch. Pressing on this button is equivalent to a mouse click.

# Speech Input/Output and Eye Tracking, Tomorrow's Apps

Let's describe how human users will interact with computers to get **simple** and **complex** work done.

| | Simple | Complex |
|---|---|---|
| **Task Description** | Scope: Small. One email, image, diagram, chart, graphic.<br><br>Effort: Single step task.<br><br>Example: Generate/revise one email, image, diagram, chart, graphic. | Scope: Big. Multi-page documents, presentations, UIs.<br><br>Effort: Multi-step chain of tasks including search, data collection, and presentation generation.<br><br>Example: Write presentation that surveys the potential investments that company X can make in its AI offering spanning infrastructure to consumer apps, and present recommended investments backed with qualitative and quantitative evidence gathered across internal/external web. |
| **User Journey of HCI** | Initial Instruction: "Give me <thing>."<br><br>Critical Review: None<br><br>Revision Instructions: Small, concrete for quick iterative revisions that are back and forth between human and computer.<br>1. "Change the tone of the email"<br>2. "Make the photo brighter"<br>3. "Make the style of the icon consistent with this attachment"<br><br>For simple revisions, human users may find it easier to make the edit themselves instead of verbalizing instructions that AI can reliably execute.<br><br>Context:<br>1. One item in workspace<br>2. Multiple items in workspace (multiple page paragraphs in document) and users must select target for revision<br><br>Deliverable: One thing | Initial Instruction: Abstract goal with success criteria.<br><br>Critical Review: Human users ask questions about how the work was done to grow trust in the computers' work, alternatives considered to assess completeness, source of data to determine credibility.<br>1. "How will the thesis of the paper change if this data is excluded because the source is iffy."<br>2. "What alternative interpretations of this data are there?"<br>3. "What are the strongest arguments against this assertion?"<br><br>Revision Instructions: Additional tasks to strengthen final deliverable such as additional research, data acquisition.<br>1. "Do analysis X. Let me know when you're done."<br>2. "Collect data X for time period Y. Chart it. Let me know when you're done."<br>3. "Revise section X so that it supports the assertion Y."<br><br>Deliverable: Multi-layered information. Presentation on top surface and information about "how task got done" and "why the computer made choices it did" underneath. |
| **HCI Mechanics** | Users will intermingle taking their own actions to revise the thing and speaking instructions for computers to execute.<br><br>Users will want to see all revision options and choose from them. Users will want to see revisions in comparison to the original. | User will ask questions. Answering some will require additional work and wait time. Answering others may require computers to look back at their work history and decision-making rationale.<br><br>Users will ask for revisions by speaking instructions.<br><br>Users will want the option of making their own fine edits with keyboard and mouse.<br><br>Users will want to see all revision options and choose from them. Users will want to see revisions in comparison to the original. |

For the most part, speaking questions and instructions is easier than typing them. The likelihood that users speak to their computers instead of typing messages is high.

A key challenge in multi-page, multi-item work products is that the computer must understand the spoken questions or instructions **in reference to a target item(s) in the body of work**. Which page of the document is the user viewing? Which chart does the revision apply to? Which image? Understanding this context will help computers more reliably understand users' questions and instructions. The likelihood that a mechanism to convey context will emerge is high.

Three patterns are possible to enable users to provide computers the context of their questions and instructions.

1. **Speech driven** - Human users verbally describe the target object.
2. **Mouse driven** - Human users select the target item(s) with mouse and then verbalize questions or instructions.
3. **Eye tracking driven** - Eye tracker always tracks users' point of focus and users verbalize questions and instructions freely.

Describing the target item is arduous for the user. Selecting with the mouse or eye tracker will be lower effort.

Eye tracking driven HCI requires the purchase of another device. However, it can help users get tasks done with less mouse manipulations.

# Future Desktop HCI

Given that the nature of the work that computers do for us will be more complex and resemble work that we're asking other humans to do today, we'll see the following new user behaviors.

1. Users will use the speech input channel to provide instructions and ask questions about how the work got done instead of typing. Speech is faster.
2. Computers will output their responses on screen because users can read faster than they can listen.
3. Users will help computers understand the context of their inquiry by pointing to onscreen objects with mice or eye trackers.
4. Users will want keyboards and mice as fallbacks.