

# Generative AI in Everyday Knowledge Work: How Employees View and Use Microsoft 365 Copilot

A simplified version of the following research paper:

*Generative AI in knowledge work: Perception, usefulness, and acceptance of Microsoft 365 Copilot.*

Carsten F. Schmidt, Sophie Petzolt, Wolfgang Beinhauer, Ingo Weber, and Stefan Langer. Preprint, 2026. <https://arxiv.org/abs/2602.18576>

## What This Study Is About

This study examines how employees in a large German non-university research organization perceive and use **Microsoft 365 Copilot**, a generative AI tool. The researchers focused on:

1. **Differences** between scientific and administrative employees' assessments of Copilot.
  2. **Changes over time** between two survey periods (T01 in late 2024 and T02 in early 2025).
  3. **Usefulness for typical knowledge-work tasks**, such as writing, information search, or learning.
- 

## Why It Matters

Generative AI is becoming standard in office environments. Yet its effects differ depending on:

- the organization's work profile,
- employee roles,
- and experience with complex vs. structured tasks.

Non-university research organizations like Fraunhofer have particularly diverse work environments involving both research and extensive administration, making them ideal testbeds.

---

## How the Study Was Done

- **Sample:** 550 employees received Copilot licenses;

- At **T01**: 106 responded (40 admin, 66 scientific).
    - At **T02**: 90 responded (39 admin, 51 scientific).
  - The survey used 7-point Likert scales (−3 to +3).
  - Because few individuals participated at both times, the study used **repeated cross-sectional** logic (not a true panel).
  - Constructs measured included:
    - Perceived usefulness
    - Perceived ease of use
    - Output quality
    - Reliability
    - Task-specific usefulness across nine activity types (e.g., analysis, text creation, networking)
- 

## What the Researchers Found

### 1. Administrative employees consistently rated Copilot more positively

At T01:

- Administration rated **usefulness** and **output quality** significantly higher than scientists.
- Both groups rated reliability and ease of use positively; differences were not significant.

At T02:

- The earlier differences largely **disappeared** because scientific employees' ratings **increased**.

### 2. Scientists' perceptions improved significantly over time

Scientific employees:

- Showed **significant increases** in perceived usefulness and ease of use between T01 and T02.
- Reported higher confidence in productivity gains, effectiveness, and workload reduction at T02.

Administrative employees:

- Started high and remained stable, showing almost no significant changes.

### 3. Copilot is most helpful for structured, text-based tasks

Both groups rated Copilot especially useful for:

- **Information search**
- **Information gathering**
- **Analysis**
- **Text creation**

These tasks consistently received positive scores, particularly in administration.

#### 4. Copilot is less helpful for creative or interpersonal tasks

Tasks with neutral or negative ratings included:

- Creating media content
- Staying up to date
- Networking with colleagues
- Learning

Both groups saw **little or no added value** here.

#### 5. Reliability and harm from incorrect outputs remained stable

- Copilot was rated as **technically reliable** at both time points.
  - Perceived harm from incorrect outputs remained *low* and unchanged.
- 

### What This Means

For organizations:

- **Role-specific training** is crucial because scientists and administrative staff have different starting points.
- Copilot should be integrated first in **structured, text-based workflows**, where benefits are clear.
- Acceptance grows primarily through **perceived usefulness**, not just ease of use.

For employees:

- Skepticism—particularly among scientific staff—declines with experience.
  - Copilot is strongest as a tool for **writing, searching, summarizing**, and other structured tasks.
- 

### Limitations

The study acknowledges several constraints:

- The sample is **not representative** of all ~32,000 employees.
- Self-selection may bias results toward more AI-interested employees.

- Assessments are **subjective**, not objective productivity data.
  - Task-related items were analyzed **descriptively**, not statistically.
  - Only early-stage adoption effects were captured.
- 

## Key Takeaways

- Administrative staff initially rated Copilot much more positively, but scientists caught up over time.
  - Copilot works best for structured, text-based knowledge work.
  - It is less useful for creative, interpersonal, or highly domain-specific research tasks.
  - Reliability is rated positively, and concerns about harmful mistakes are low.
  - Sustainable AI adoption requires training, governance, and context-specific support.
- 

This simplified version was created using an AI agent