The impact of computer-aided detection systems on typo foraging

Emily M. Heffernan, Benjamin Wolfe, Anna Kosovicheva
Department of Psychology, University of Toronto Mississauga

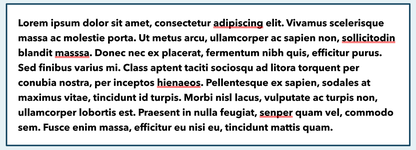
UNIVERSITY OF TORONTO

APPLY LAB

Introduction

Our performance in visual search tasks is imperfect—we often miss clearly visible targets.

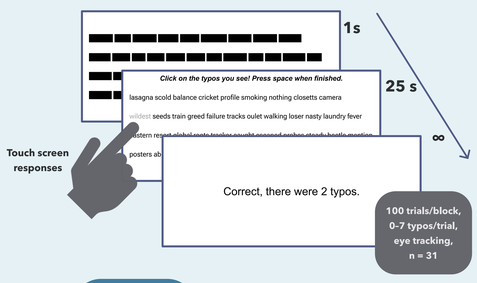
Computer-aided detection (CAD) tools are an increasingly common means of directing our attention to areas of interest.

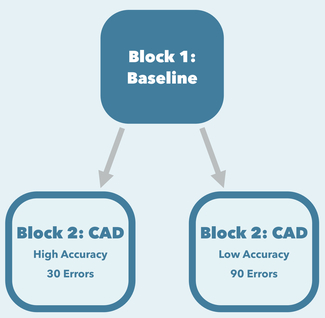


How does CAD influence attention and performance in a typo foraging task?

Methods

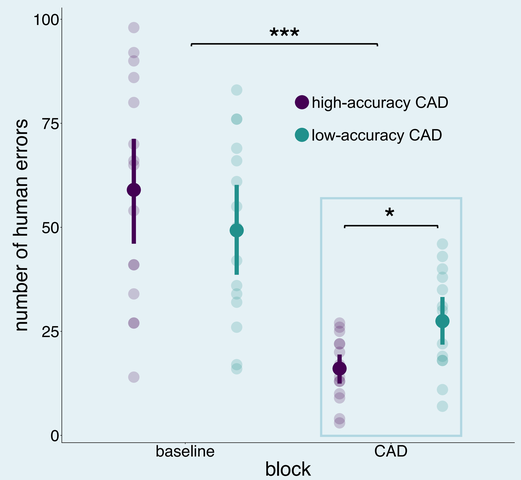
Typo Foraging Task

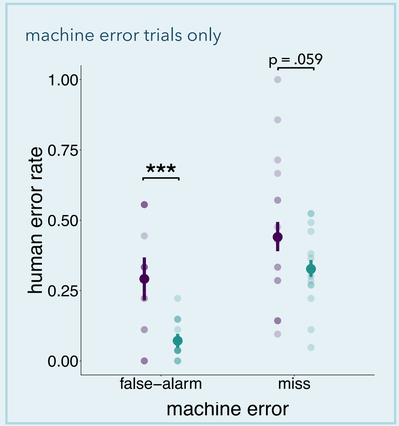




CAD errors throughout block	
False Alarm High-Acc: 4 Low-Acc: 12	liner worse skill wealthy mellow plains fleece easiest glitter proved
Miss High-Acc: 16 Low-Acc: 48	liner worse skill wealthy melllow plains fleece easiest glitter proved
False Alarm + Miss High-Acc: 10 Low-Acc: 30	liner worse skill wealthy melllow plains fleece easiest glitter proved

How does CAD impact the number of human errors?



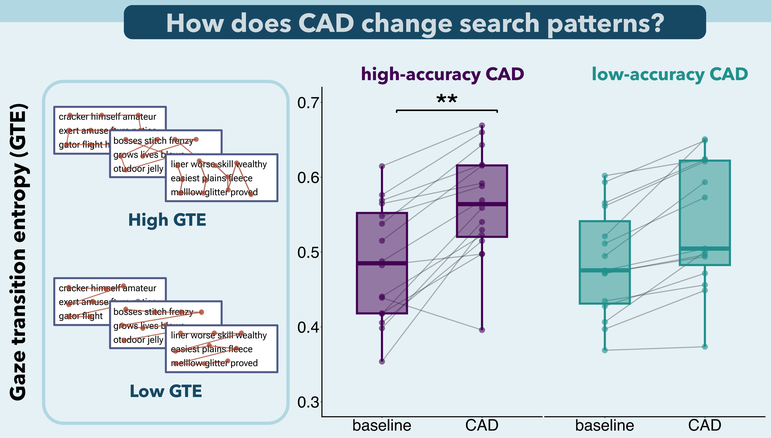


Overall, CAD reduces the number of errors

Participants with the **high-accuracy** CAD do better than those with the **low-accuracy** CAD

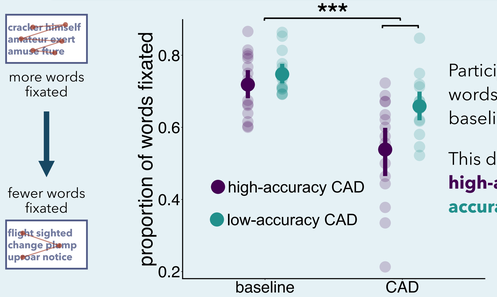
When CAD is incorrect, participants in the **low-accuracy** condition have lower error rates than participants in the **high-accuracy condition**

How does CAD change search patterns?



Scanning patterns become significantly more random only for participants in the **high-accuracy** CAD condition

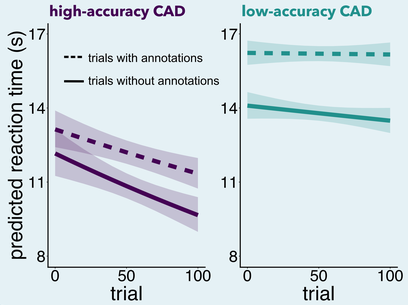
How thoroughly are participants looking?



Participants fixate fewer words in the CAD relative to baseline block

This drop is larger in the **high-accuracy** versus **low-accuracy** CAD condition.

And for how long?



During the CAD block:

Participants with the **high-accuracy** CAD search for less time than those with the **low-accuracy** CAD.

They also get faster over time.

In general, participants spend more time scanning trials with annotations versus trials without.

Takeaways

CAD tools augment typo detection, but this benefit is not without attentional consequences.

When a system seems reliable, we detect typos in a faster, but less thorough, manner.

Consequently, we more readily overlook even blatant machine errors, like false alarms.

What's next?

Which factors lead to individual differences in over-reliance on automated tools?

How quickly does trust in automation develop or disappear, and can such changes be reversed?

Can input data (e.g., mouse movements) be used as a live indicator of over-reliance?

Acknowledgements

Thank you to APPLY Lab members for help and feedback! This work was supported by an SSHR Insight Grant to BW and AK.

Poster pdf here