

# We Write Our Research Papers in WYSIWYM.

## Why Do We Tag Our PDFs in WYSIWYG?

<Peya Mowar, Aaron Steinfeld, Jeffrey Bigham>

Carnegie Mellon University 

<WYSIWYM>

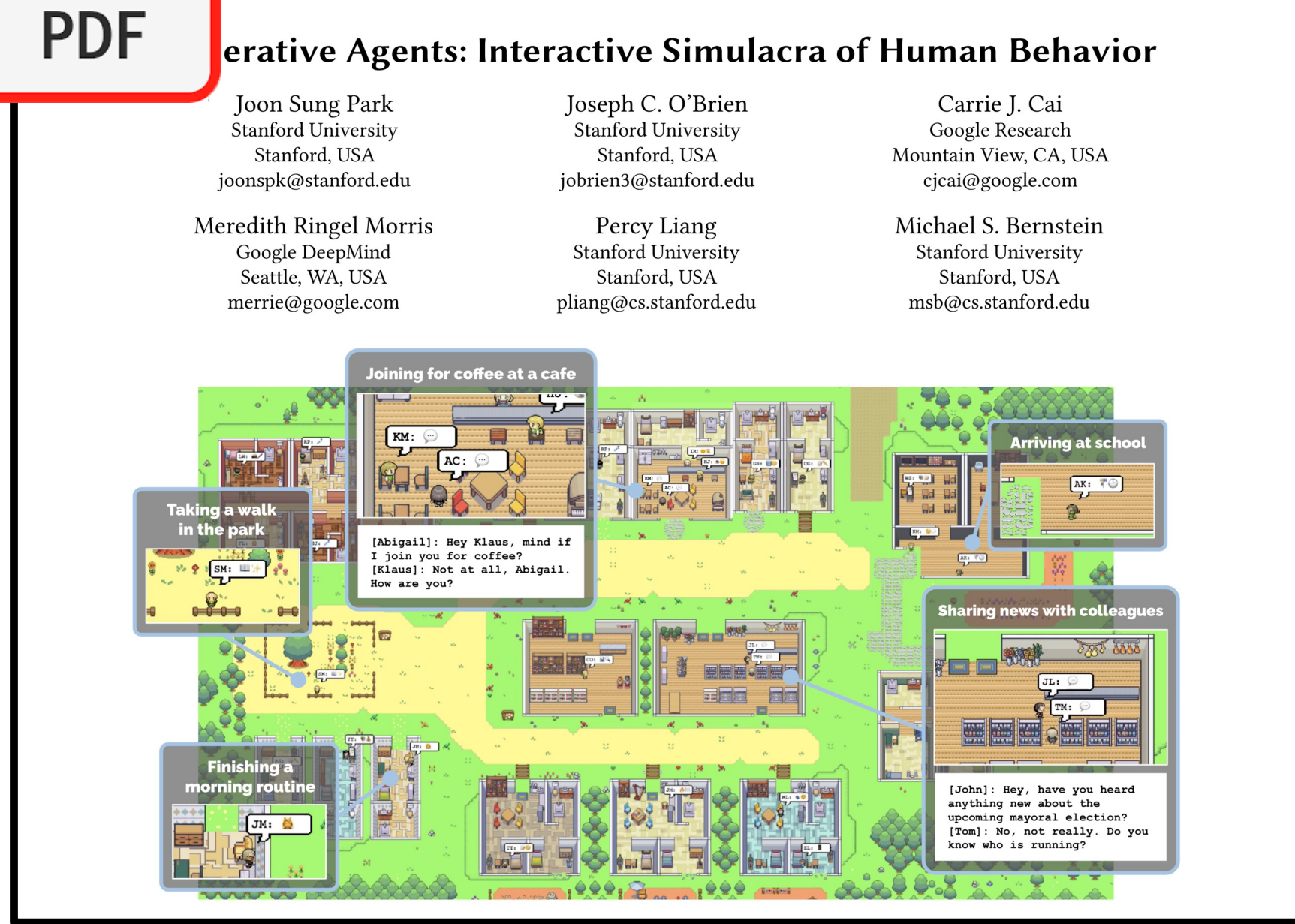
What You See Is What You Mean

### Authoring Tools

L<sup>A</sup>T<sub>E</sub>X

```
\title{Generative Agents}{Generative Agents: Interactive Simulacra of Human Behavior}
\author{Joon Sung Park}
\affiliation{%
  \institution{Stanford University}
  \city{Stanford}
  \country{USA}}
\email{joonspk@stanford.edu}
```

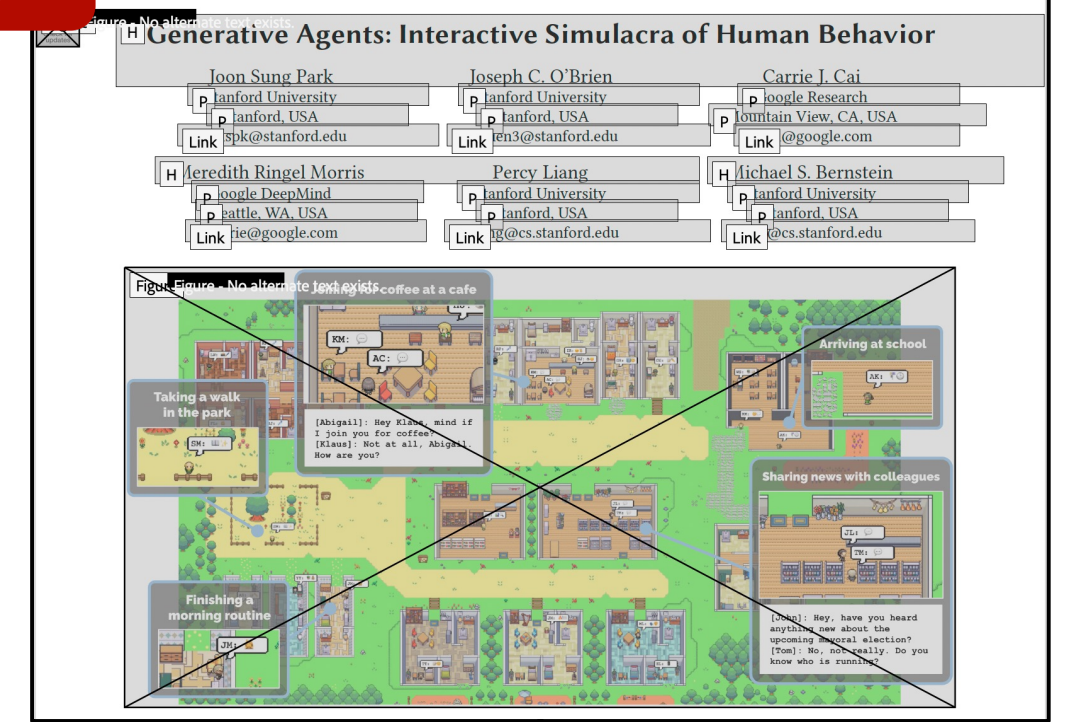
Contain **Semantic** Meaning



<WYSIWYG>

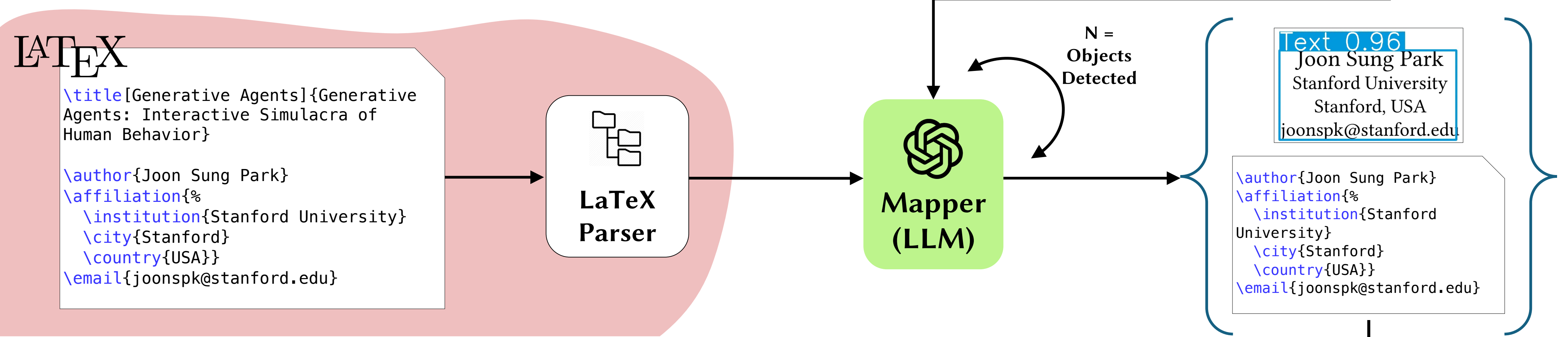
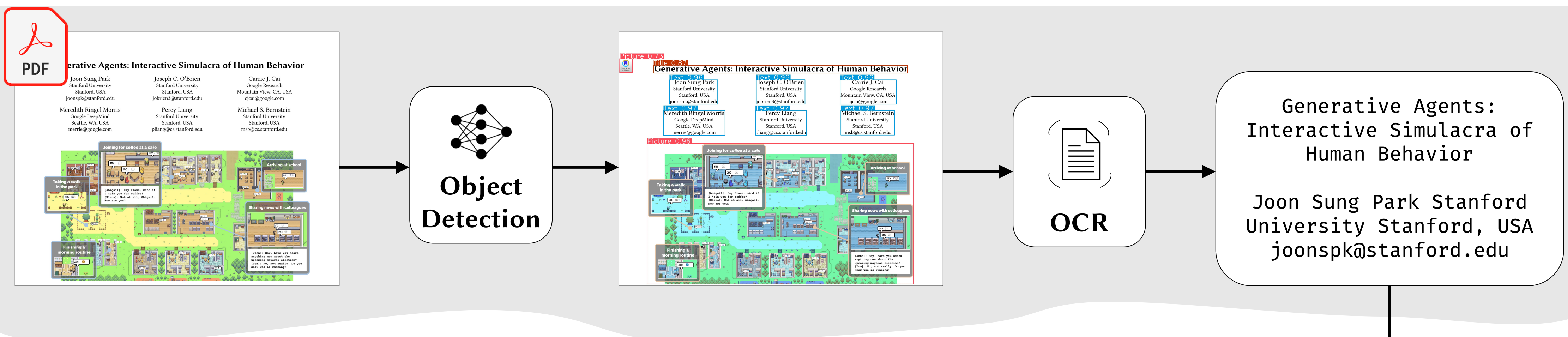
What You See Is What You Get

### PDF Remediation Tools



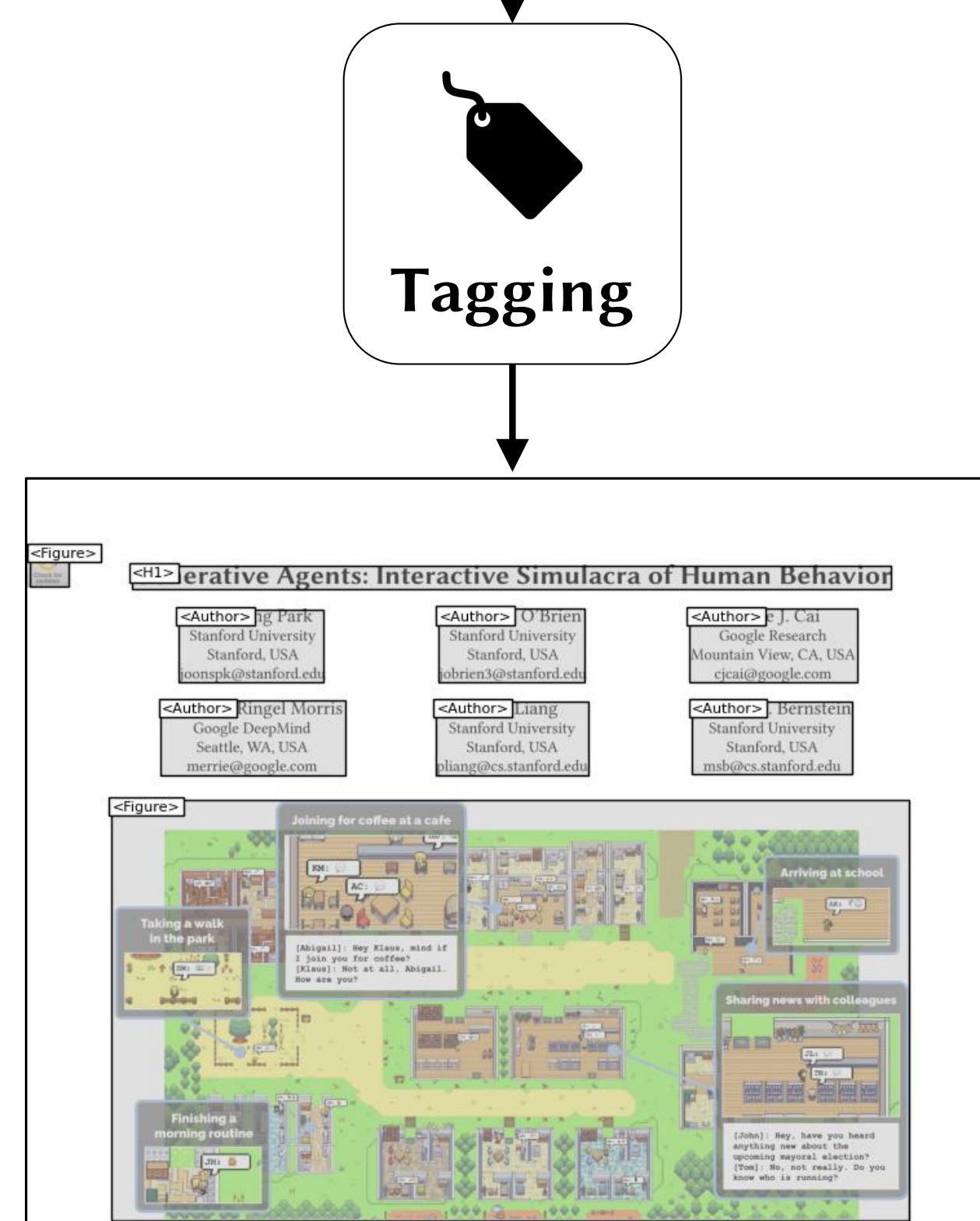
Consider **Visual** Rendering

**How can we preserve the semantic metadata naturally generated during authoring but typically discarded during visual PDF rendering?**



Tagging Results on N = 10 Research Paper PDFs published at ACM ASSETS (2019-2024)

Criteria	Adobe Acrobat Auto-Tags [%]	What Authors Submit to ASSETS [%]	WYSIWYM Tags (Ours) [%]
All Content Tagged	100.0	100.0	100.0
Reading Order	89.66	95.42	<b>*96.26*</b>
Headings Tagged	90.47	94.78	82.68
Headings Tagged + Level	85.28	<b>*94.35*</b>	81.82
Tables Tagged	68.18	50.0	<b>*95.0*</b>
Lists Tagged	60.34	85.36	75.61
Figures Tagged	52.24	<b>*100.0*</b>	84.0
Captions Tagged	0.0	6.45	<b>*95.65*</b>
Average Score	71.79	78.21	<b>*88.75*</b>



**New!** Even better results, ask me more!