**Key Findings**

**Human+AI team** can exploit complementary strengths of both humans and machines and surpass either.

Statistically significant improvements in work quality; improved worker productivity.

**Labor Specialization: AI** does better on lower-complexity work; humans better in more complex regimes.

Evidence for J.C.R. Licklider’s hypothesis that **human-machine symbiosis** can harness the “cognitive” capabilities of both humans and machines optimally.

---

**Understanding Human-AI Work Collaboration using a Randomized Field Study**

Abhinav Maurya\(^4\), Sunder Kekre\(^5\), Rahul Telang\(^6\)

\(^4\)Heinz College of Information Systems and Public Policy, \(^5\)Machine Learning Department, \(^6\)Tepper School of Business, Carnegie Mellon University

---

**Designing a Study to Compare AI, Human, and AI+Human Team**

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Human Workers</th>
<th>Annotations</th>
<th>Control Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Data</td>
<td>Human Learning NE Suggestions</td>
<td>AI Evaluation</td>
<td>Groundtruth Annotations</td>
</tr>
<tr>
<td><strong>Control (Human)</strong></td>
<td><strong>Treatment (AI+Human)</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Running The 3-Day Experiment on Named Entity Annotation**

- **Control group** (n participants who carried out the same tasks overall)
- **Treatment group** (n participants who carried out the same tasks as control group)

---

**Error Decomposition Trees for AI, Control, and Treatment**

- AI Predictions
- Control (Human)
- Treatment (Human+AI)

---

**Comparing AI, Control, and Treatment**

**Work Quality**

<table>
<thead>
<tr>
<th>Metric</th>
<th>AI</th>
<th>Control</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>0.952</td>
<td>0.951</td>
<td>0.951</td>
</tr>
<tr>
<td>FN</td>
<td>0.027</td>
<td>0.027</td>
<td>0.027</td>
</tr>
<tr>
<td>Precision</td>
<td>0.734</td>
<td>0.734</td>
<td>0.734</td>
</tr>
<tr>
<td>Recall</td>
<td>0.868</td>
<td>0.868</td>
<td>0.868</td>
</tr>
<tr>
<td>F1-Score</td>
<td>0.798</td>
<td>0.798</td>
<td>0.798</td>
</tr>
</tbody>
</table>

---

**Conclusions and Future Work**

- Studied mediating factors of Human-AI collaboration in NLP data annotation services.
- Potential generalization beyond NLP e.g. computer vision where work complexity in object detection might be measured by bounding box complexity.
- **Theory** for what team construction regimes are better.
- Investigating **pathological regimes** of Human-AI team dynamics.
- **Engendering trust** in AI agents in Human-AI teams.
- Multistage work and cooperation in Human-AI teams.

---

Working Paper