

## Appendix 1: Software Components used in the MARDY Annotation Environment

- XML-repository: java-based backend for NLP enriched newspaper articles (based on UIMA framework)
- ElasticStack: native storage for JSON annotations, predictions, article data; Web-Fronted (Kibana) allows direct inspection of all data
- PostgreSQL: additional data storage (Elastic is not recommended as productive backend), user management (e.g., article assignment), annotation management (article assignment)
- WebApp: Frontend: AngularJS+AnnotatorJS+Bootstrap; Backend: native Java Servlets + JSP pages

## Appendix 2: Regression Analysis of the Annotation Experiment

Table A1 contains the regression analysis to test whether the AI suggestions significantly speed up the annotation process. The first two models are OLS-regressions controlling for the number of claims (model 1), article length in tokens, and whether the participant was a student or a senior researcher (model 2). In both models, all coefficients, except for the senior researcher dummy variable, are statistically significant ( $p < 0.05$ , we used robust standard errors for all models). The last two models are fixed-effects regressions accounting for omitted variable bias both on the level of participants (model 3) and over the four distinct rounds by introducing a time trend (model 4). As soon as we introduce fixed effects for annotators, the coefficient for the AI suggestions is no longer significant. Based on this, we conclude that there might be omitted variables (such as intelligence) that are heterogeneous across annotators and bias our results. This holds true even if we control for learning effects. Note that two of the annotators exceeded the time limit and did not finish their assigned set of articles (without AI suggestions). In order to balance the data and our models, we excluded two articles in round 2 from the regression analysis.

**Table A1:** Regression analysis of annotation speed due to AI suggestions

	<i>Dependent variable: seconds</i>			
	<i>OLS</i>		<i>panel linear</i>	
	(1)	(2)	(3)	(4)
AI suggestions (yes)	-53.126** (22.566)	-39.273** (18.683)	-43.272 (33.598)	-42.258 (32.402)
number of claims	35.464*** (3.152)	22.930*** (1.996)	23.823*** (2.341)	22.775*** (1.857)
number of tokens		0.253*** (0.045)	0.243*** (0.044)	0.255*** (0.045)
senior researcher (yes)		-26.799 (18.142)		
time trend (R1-R4)				-30.475*** (4.306)
Constant	92.254*** (25.815)	58.115** (25.299)		
Observations	228	228	228	228
R <sup>2</sup>	0.710	0.783	0.796	0.811
Adjusted R <sup>2</sup>	0.707	0.779	0.788	0.803
Residual Std. Error	157.458 (df = 225)	136.934 (df = 223)		
F Statistic	275.363*** (df = 2; 225)	200.671*** (df = 4; 223)	284.217*** (df = 3; 219)	233.187*** (df = 4; 218)

Note: \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$