

# users	1,000,000	# impressions	15,777,377
# news	161,013	# news click behaviors	24,155,470
avg. title len.	11.52	# training samples	2,186,683
# validation samples	365,200	# test samples	2,341,619

Extended Data Table 2. Statistics of the MIND dataset. The average title length means the average number of tokens in the news title. Each sample is composed of a user’s historical clicked news articles and the click/non-click behaviors on a candidate news article.