TABLE V

CHARACTERIZATION OF (44) EXPERIMENTS THAT EARNED ACM ARTIFACT BADGES. CELLS WITH A SINGLE LABEL MEAN THE ARTIFACT DOES NOT IMPROVE THE INFORMATION PROVIDED IN THE PAPER. CELLS WITH AN ARROW MEAN THAT THE INFORMATION PROVIDED IN THE ARTIFACT IMPROVES THE INFORMATION PROVIDED IN THE PAPER. LIGHT GRAY BACKGROUND REPRESENTS AN INCONSISTENCY BETWEEN THE CONTENT OF THE PAPER AND THE ARTIFACT. DARK GRAY BACKGROUND REPRESENTS INCONSISTENCIES WITHIN THE ELEMENTS IN THE ARTIFACT.

		Hyp.	Variables identification					Operationalization		Design		Objects sel.	Analysis & interpr.		Val. eval.
		Res. hyp.	Model hyper.	Model params.	DL alg.	Training hyper.	Training data	Factors & treat.	Resp. vars.	Choice design	Instrum.	Test set chars.	Descr. stats.	Infer. stats.	Validity threats
[AP5]	E1	FA	PA→FA	M→FA	PA	PA	FA	PA→PA	FA	M	PA→PA	M	M	M	M
	E2	FA	PA→FA	M→FA	PA	PA	FA	PA→PA	FA	PA	PA→PA	M	M	M	M
	E3	FA	PA→FA	M→FA	PA	PA	FA	PA→PA	FA	PA	PA	M	M	M	M
	E4	FA	PA→FA	M→FA	PA	PA	FA	PA	FA	PA	PA	M	M	M	M
[AP8]	E1	FA	PA→FA	FA	FA	PA→FA	FA	PA→PA	FA	M	PA	PA	M	M	PA
	E2	FA	PA→FA	FA	FA	PA→FA	FA	PA→PA	FA	M	PA	PA	M	M	PA
	E3	FA	PA→FA	FA	FA	PA→FA	FA	PA→PA	FA	M	PA	PA	M	M	PA
	E4	FA	PA→FA	FA	FA	PA→FA	FA	PA→PA	FA	M	PA	PA	M	M	PA
[AP10]	E1	M	PA→FA	M	PA→FA	PA→FA	FA	PA→PA	M→FA	PA	PA	PA	M	M	PA
	E2	FA	PA→FA	M	PA→FA	PA→FA	FA	PA→PA	FA	PA	PA	PA	FA	FA	PA
	E3	FA	PA→FA	M	PA→FA	PA→FA	FA	PA→PA	FA	PA	PA	PA	FA	M	PA
	E4	FA	PA	M	PA	PA	FA	FA	FA	PA	PA	PA	FA	M	PA
[AP15]	E1	FA	PA→FA	M	PA→FA	PA→FA	FA	PA→PA	FA	PA	PA	M	PA	M	PA
	E2	FA	PA→FA	M	PA→FA	PA→FA	FA	PA→PA	FA	PA	PA	M	PA	M	PA
	E3	FA	PA→FA	M	PA→FA	PA→FA	FA	PA→PA	FA	PA	PA	M	PA	M	PA
[AP29]	E1	FA	PA→FA	M	PA→FA	PA→FA	FA	PA	PA→FA	PA	PA	PA	PA	M	M
[AP36]	E1	FA	PA→FA	M	PA→FA	PA→FA	FA→FA	PA→PA	FA	PA	FA	PA	PA	M	M
	E2	FA	PA→FA	M	PA→FA	PA→FA	FA→FA	PA	FA	M	FA	PA	M	M	M
	E3	FA	PA→FA	M	PA→FA	PA→FA	FA→FA	FA	M→FA	M	FA	PA	M	M	M
	E4	FA	PA→FA	M	PA→FA	PA→FA	FA→FA	PA→PA	FA	PA	PA→FA	PA	M	M	M
[AP39]	E1	FA	FA	M	FA	FA	FA	PA	FA	PA	PA	FA	M	M	PA
	E2	FA	FA	M	FA	FA	FA	PA	FA	PA	PA	FA	PA	M	PA
	E3	FA	FA	M	FA	FA	FA	PA	FA	PA	PA	FA	PA	M	PA
	E4	FA	FA	M	FA	FA	FA	PA	FA	PA	PA	FA	PA	M	PA
	E5	FA	FA	M	FA	FA	FA	PA	FA	PA	PA	FA	PA	M	PA
	E6	FA	FA	M	FA	FA	FA	PA	FA	M	PA	FA	M	M	PA
	E7	FA	FA	M	FA	FA	FA	PA	FA	M	PA	FA	M	M	PA
	E8	FA	FA	M	FA	FA	FA	PA	FA	M	PA	FA	PA	M	PA
	E9	M	M→FA	M	M→FA	M→FA	M→FA	M→FA	M→PA	M	M→PA	M→FA	M→PA	M	M
	E10	M	M→FA	M	M→FA	M→FA	M→FA	M→FA	M→FA	M	M→PA	M→FA	M	M	M
	E11	M	M→FA	M	M→FA	M→FA	M→FA	M→PA	M→FA	M	M→PA	M→FA	M	M	M
	E12	FA	FA	M	FA	FA	FA	FA	FA	FA	FA	FA	FA	FA	FA
[AP40]	E1	FA	PA→FA	M	FA	PA→FA	PA→FA	PA→PA	FA	PA	PA	M	M	M	PA
	E2	FA	PA→FA	M	FA	PA→FA	PA→FA	PA→PA	FA	PA	PA	M	M	M	PA
	E3	FA	PA→FA	M	FA	PA→FA	PA	PA	FA	PA	PA	M	M	M	PA
	E4	FA	PA→FA	M	FA	PA→FA	PA→FA	PA→PA	FA	PA	PA	M	M	M	PA
	E5	M	PA→FA	M	FA	PA→FA	PA	PA	FA	PA	PA	M	FA	M	PA
[AP41]	E1	M	PA	M	PA	PA	M	FA	M	M	M	M	M	M	PA
	E2	FA	PA→FA	M	PA→FA	PA→FA	FA	PA→PA	FA	PA	PA	M	M	M	PA
	E3	FA	PA→FA	M	PA→FA	PA→FA	FA	PA→PA	FA	M	PA	M	M	M	PA
	E4	FA	PA→FA	M	PA→FA	PA→FA	FA	PA→PA	FA	PA	PA	M	M	M	PA
	E5	FA	PA→FA	M	PA→FA	PA→FA	FA	PA→PA	FA	M	PA	M	M	M	PA
	E6	M	PA→FA	M	PA→FA	PA→FA	FA	PA→PA	FA	M	PA	M	M	M	PA
	E7	M	PA→FA	M	PA→FA	PA→FA	FA	FA	FA	PA	PA	M	M	M	PA

AP5:

		E1	E2	E3	E4
Hypotheses formulation	Research hypotheses	-	-	-	-
Variables identification	Model hyperparameters	Discrepancy between paper and file “model.h5” provided	Same as previous	Same as previous	Same as previous
	Model parameters	File .h5 provided	Same as previous	Same as previous	Same as previous
	DL algorithm	Training code not provided	Same as previous	Same as previous	Same as previous
	Training hyperparameters	Training code not provided	Same as previous	Same as previous	Same as previous
	Training data	Provided (dropbox link). Matches paper	Same as previous	Same as previous	Same as previous
Operationalization	Factors and treatments	Discrepant model hyperparameters. Training code not provided	Same as previous for NL2Type. DeepTyper DNN is taken from the authors and further trained, but not included in artifact	Same as E1	-
	Response variables	-	-	-	-
Design	Choice of design	-	-	-	-
	Instrumentation	Test set provided (dropbox link)	Same as previous	-	-
Objects selection	Test set characteristics	-	-	-	-
Analysis & interpretation	Descriptive statistics	-	-	-	-
	Inferential statistics	-	-	-	-
Validity evaluation	Validity threats	-	-	-	-

AP8:

		E1	E2	E3	E4
Hypotheses formulation	Research hypotheses	-	-	-	-
Variables identification	Model hyperparameters	Inconsistencies with code	Same as previous	Same as previous	Same as previous
	Model parameters	Missing. Inconsistency with paper (says it is provided)	Same as previous	Same as previous	Same as previous
	DL algorithm	Inconsistencies with code	Same as previous	Same as previous	Same as previous
	Training hyperparameters	Inconsistencies with code	Same as previous	Same as previous	Same as previous
	Training data	-	-	-	-
Operationalization	Factors and treatments	Improved, but inconsistencies might affect	Improved, but inconsistencies might affect	Improved, but inconsistencies might affect	Improved, but inconsistencies might affect
	Response variables	-	-	-	-
Design	Choice of design	-	-	-	-
	Instrumentation	-	-	-	-
Objects selection	Test set characteristics	-	-	-	-
Analysis & interpretation	Descriptive statistics	-	-	-	-
	Inferential statistics	-	-	-	-
Validity evaluation	Validity threats	-	-	-	-

AP10:

E4 is missing from the code

		E1	E2	E3	E4
Hypotheses formulation	Research hypotheses	-	-	-	-
Variables identification	Model hyperparameters	Matches and completes	Matches and completes	Matches and completes	-
	Model parameters	-	-	-	-
	DL algorithm	Matches and completes	Matches and completes	Matches and completes	-
	Training hyperparameters	Matches and completes	Matches and completes	Matches and completes	-
	Training data	Matches	Matches	Matches	-
Operationalization	Factors and treatments	Completes	Same as previous	Same as previous	-
	Response variables	Testing error (completes)	-	-	-
Design	Choice of design	The 30 times can be changed when invoking the function (not really inconsistency)	The 30 times can be changed when invoking the function (not really inconsistency)	The 30 times can be changed when invoking the function (not really inconsistency)	-
	Instrumentation	Matches	Matches	Matches	-
Objects selection	Test set characteristics	-	-	-	-
Analysis & interpretation	Descriptive statistics	-	-	-	-
	Inferential statistics	-	-	-	-
Validity evaluation	Validity threats	-	-	-	-

AP15:

		E1	E2	E3
Hypotheses formulation	Research hypotheses	-	-	-
Variables identification	Model hyperparameters	Now fully addressed. No contradictory info.	Now fully addressed. No contradictory info.	Now fully addressed. No contradictory info.
	Model parameters	-	-	-
	DL algorithm	Contradictory info (optimization and loss function)	Contradictory info (optimization and loss function)	Contradictory info (optimization and loss function)
	Training hyperparameters	Contradictory info (epochs) in different files. No reference to grid search	Contradictory info (epochs) in different files. No reference to grid search	Contradictory info (epochs) in different files. No reference to grid search
	Training data	No train_d and train_t (splitted data). Inconsistency with paper	No train_d and train_t (splitted data). Inconsistency with paper	No train_d and train_t (splitted data). Inconsistency with paper
Operationalization	Factors and treatments	Issues due to previous contradictions	Issues due to previous contradictions	Issues due to previous contradictions
	Response variables	-	-	-
Design	Choice of design	-	-	-
	Instrumentation	-	-	-
Objects selection	Test set characteristics	-	-	-
Analysis & interpretation	Descriptive statistics	-	-	-
	Inferential statistics	-	-	-
Validity evaluation	Validity threats	-	-	-

AP29:

		E1
Hypotheses formulation	Research hypotheses	-
Variables identification	Model hyperparameters	Discrepancy in the number of neurons per layer. Now complete
	Model parameters	-
	DL algorithm	Discrepancy with loss function. Now complete.
	Training hyperparameters	Now complete
	Training data	Not provided. Must be requested from authors. Paper mentions that they share code and data
Operationalization	Factors and treatments	Not all treatments are included. Discrepancy
	Response variables	Not all of them. Discrepancy
Design	Choice of design	-
	Instrumentation	-
Objects selection	Test set characteristics	-
Analysis & interpretation	Descriptive statistics	-
	Inferential statistics	-
Validity evaluation	Validity threats	-

AP36:

		E1	E2	E3	E4
Hypotheses formulation	Research hypotheses	-	-	-	-
Variables identification	Model hyperparameters	Now complete, but there are commented lines in the decoder	Same as previous	Same as previous	Same as previous
	Model parameters	Model is saved and reloaded to avoid crashes, but not stored	Same as previous	Same as previous	Same as previous
	DL algorithm	Now complete	Same as previous	Same as previous	Same as previous
	Training hyperparameters	Now complete. But there are discrepancies	Same as previous	Same as previous	Same as previous
	Training data	Names do not match the ones in the paper	Same as previous	Same as previous	Same as previous
Operationalization	Factors and treatments	Inconsistencies due to variables above	Same as previous	Same as previous	Same as previous
	Response variables	-	-	Completes	-
Design	Choice of design	-	-	-	-
	Instrumentation	Names of datasets do not match the ones in the paper	Same as previous	Same as previous	Same as previous
Objects selection	Test set characteristics	-	-	-	-
Analysis & interpretation	Descriptive statistics	-	-	-	-
	Inferential statistics	-	-	-	-
Validity evaluation	Validity threats	-	-	-	-

AP39:

		E1	E2	E3	E4
Hypotheses formulation	Research hypotheses	-	-	-	-
Variables identification	Model hyperparameters	Possible inconsistencies between supplementary material and code	Same as previous	Same as previous	Same as previous
	Model parameters	Pre-trained available in paper (but link is broken)	Same as previous	Same as previous	Same as previous
	DL algorithm	Yes	Same as previous	Same as previous	Same as previous
	Training hyperparameters	Yes	Same as previous	Same as previous	Same as previous
	Training data	Match neither paper nor supplementary material.	Same as previous	Same as previous	Same as previous
Operationalization	Factors and treatments	Possible inconsistencies due to model hyperparam	Same as previous	Same as previous	Same as previous
	Response variables	-	-	-	-
Design	Choice of design	-	-	-	-
	Instrumentation	Datasets match neither paper nor supplementary material	Same as previous	Same as previous	Same as previous
Objects selection	Test set characteristics	Ok	Ok	Ok	Ok
Analysis & interpretation	Descriptive statistics	-	-	-	-
	Inferential statistics	-	-	-	-
Validity evaluation	Validity threats	-	-	-	-

		E5	E6	E7	E8
Hypotheses formulation	Research hypotheses	-	-	-	-
Variables identification	Model hyperparameters	Same as previous	Same as previous	Same as previous	Same as previous
	Model parameters	Same as previous	Same as previous	Same as previous	Same as previous
	DL algorithm	Same as previous	Same as previous	Same as previous	Same as previous
	Training hyperparameters	Same as previous	Same as previous	Same as previous	Same as previous
	Training data	Same as previous	Same as previous	Same as previous	Same as previous
Operationalization	Factors and treatments	Same as previous. Artifact does not seem to contain ablation study	Same as previous	Same as previous	Same as previous
	Response variables	-	-	-	-
Design	Choice of design	-	-	-	-
	Instrumentation	Same as previous	Same as previous	Same as previous	Same as previous
Objects selection	Test set characteristics	Ok	Ok	Ok	Ok
Analysis & interpretation	Descriptive statistics	-	-	-	-
	Inferential statistics	-	-	-	-
Validity evaluation	Validity threats	-	-	-	-

These are the ones that appear in the supplementary material. The last one (E13) is the only one specifically referenced in the paper.

		E9	**E10**	**E11**	**E12**
Hypotheses formulation	Research hypotheses	-	-	-	-
Variables identification	Model hyperparameters	Same as previous	Same as previous	Same as previous	Same as previous
	Model parameters	Same as previous	Same as previous	Same as previous	Same as previous
	DL algorithm	Same as previous	Same as previous	Same as previous	Same as previous
	Training hyperparameters	Same as previous	Same as previous	Same as previous	Same as previous
	Training data	Same as previous	Same as previous	Same as previous	Same as previous
Operationalization	Factors and treatments	Predicted type	Test set	Model type (Stateformer, Debin)	Numerical values embedding, number of layers, layers dimensions,
	Response variables	Accuracy (just mentioned)	F1	F1	Training loss
Design	Choice of design	No	No	No	No
	Instrumentation	Same as others	Same as others	Same as others	Same as others
Objects selection	Test set characteristics	Yes	Yes	Yes	Yes
Analysis & interpretation	Descriptive statistics	Average only	No	No	No
	Inferential statistics	No	No	No	No
Validity evaluation	Validity threats	No	No	No	No

AP40:

		E1	E2	E3	E4	E5
Hypotheses formulation	Research hypotheses	-	-	-	-	-
Variables identification	Model hyperparameters	Now appear. Do not contradict paper because in paper they are not described	Same as previous	Same as previous	Same as previous	Same as previous
	Model parameters	-	-	-	-	-
	DL algorithm	AdamW also appears in code	Same as previous	Same as previous	Same as previous	Same as previous
	Training hyperparameters	Now complete and do not contradict paper	Same as previous	Same as previous	Same as previous	Same as previous
	Training data	In paper it is not specifically linked, but appears in RP	Same as previous	-	Same as E1-E2	-
Operationalization	Factors and treatments	Possible inconsistencies due to DL algorithm	Same as previous	Treatments are different subsets of training set. They do not appear in artifact. Does not improve	Same as E1-E2	Treatments are different subsets of training set. They do not appear in artifact. Does not improve
	Response variables	-	-	-	-	-
Design	Choice of design	-	-	-	-	-
	Instrumentation	-	-	-	-	-
Objects selection	Test set characteristics	-	-	-	-	-
Analysis & interpretation	Descriptive statistics	-	-	-	-	-
	Inferential statistics	-	-	-	-	-
Validity evaluation	Validity threats	-	-	-	-	-

AP41:

		E1	E2	E3	E4
Hypotheses formulation	Research hypotheses	-	-	-	-
Variables identification	Model hyperparameters	-	Yes. There is a configuration file (data_util/config.py)	Same as previous	Same as previous
	Model parameters	-	-	-	-
	DL algorithm	-	Configuration file	Same as previous	Same as previous
	Training hyperparameters	-	Configuration file. There are two (commented) values for epochs. Inconsistency	Same as previous	Same as previous
	Training data	-	Linked	Linked	Linked
Operationalization	Factors and treatments	-	Possible inconsistencies due to training hyperparameters	Same as previous	Same as previous
	Response variables	-	-	-	-
Design	Choice of design	-	-	-	-
	Instrumentation	-	-	-	-
Objects selection	Test set characteristics	-	-	-	-
Analysis & interpretation	Descriptive statistics	-	-	-	-
	Inferential statistics	-	-	-	-
Validity evaluation	Validity threats	-	-	-	-

		E5	E6	E7
Hypotheses formulation	Research hypotheses	-	-	-
Variables identification	Model hyperparameters	Same as previous	Same as previous	Same as previous
	Model parameters	-	-	-
	DL algorithm	Same as previous	Same as previous	Same as previous
	Training hyperparameters	Same as previous	Same as previous	Same as previous
	Training data	Linked	Linked	Linked
Operationalization	Factors and treatments	Same as previous	Same as previous	-
	Response variables	-	-	-
Design	Choice of design	-	-	-
	Instrumentation	-	-	-
Objects selection	Test set characteristics	-	-	-
Analysis & interpretation	Descriptive statistics	-	-	-
	Inferential statistics	-	-	-
Validity evaluation	Validity threats	-	-	-