ONLINE APPENDIX

Supplemental Tables to "Using supervised machine learning to scale human-coded data: An illustration in the board leadership context"

Joseph S. Harrison
Department of Management and Leadership
Neeley School of Business
Texas Christian University
2900 Lubbock Ave, Fort Worth, TX 76109
j.s.harrison@tcu.edu

Matthew A. Josefy
Department of Management and Entrepreneurship
Kelley School of Business
Indiana University
1309 East Tenth Street, Bloomington, IN 47405
mjosefy@iu.edu

Matias Kalm
Department of Management
Tilburg School of Economics and Management
Warandelaan 2, 5037 AB Tilburg, Netherlands
m.k.j.kalm@tilburguniversity.edu

Ryan Krause
Department of Management and Leadership
Neeley School of Business
Texas Christian University
2900 Lubbock Ave, Fort Worth, TX 76109
r.krause@tcu.edu

Contents

TABLE S1 Definitions of board chair orientations

Construct	Definition
Control orientation	"[The control] orientation is based on the belief that a separate individual should act as board chair so that the chair can monitor, oversee, and if necessary, discipline the CEO. Boards exhibiting this orientation will often use the words 'oversight' and 'independence' or explain that a separate board chair facilitates holding the CEO and management accountable, evaluating the management, or introducing greater objectivity and integrity in board decision-making. The underpinning assumption of this orientation is that the role of the board chair is to help the board control the CEO. (Krause, 2017: Appendix S1)
Collaboration orientation	"[The collaboration] orientation is based on the belief that the board chair's role is to advise and guide the CEO, as well as to help the CEO perform his or her job by reducing the demands on the CEO's time. This orientation often involves distinguishing the role of the CEO (day-to-day leadership of the firm) from that of the board chair (leading the board and providing broad strategic direction), and suggests that by filling different roles, the CEO and board chair can specialize in their responsibilities. Boards exhibiting this orientation will often note that a separate board chair enables the CEO to devote all his/her attention to managing the firm, improves communication between the board and management, or helps the board to provide advice and guidance to the CEO. The underpinning assumption of this orientation is that the role of the board chair is to help the board collaborate with the CEO." (Krause, 2017: Appendix S1)

TABLE S2 Feature extraction methods

Model	General Description
Count Vector (CV)	Generates a feature matrix with as many unique terms (i.e., individual words, word stems, or phrases) as exist in the text corpus. Alternatively, the total number of features may be restricted to *n* prior to modeling, where 1 to *n* reflect the *n* most frequent terms in the corpus. For a given document, each feature is equal to the number of occurrences of the term in the document.
Term Frequency - Inverse Document Frequency (TF-IDF) Vector	Generates the same size feature matrix as CV, but rather than simple counts, uses a weighting scheme that multiplies a term frequency by its inverse document frequency. Weighting is intended to provide an indication of the importance of the term in the corpus.
Word2Vec (W2V) Embedding	Developed by Mikolov at al. (2013). Generates a multidimensional vector space with each unique word in the corpus being assigned a distinct vector in the space. Each vector is positioned so that semantically similar words are located close to one another in the space. During training, models learn to predict words using one of two models: • *Skip-gram* uses a target word to predict the *n* words before and after the target word. • *Continuous bag of words (CBOW)* uses the *n* words before and after the target word to predict the target word.
Doc2Vec (D2V) Embedding	Developed by Le and Mikolov (2014) as an extension of W2V. Generates a multidimensional vector space like W2V, but also includes a vector for each document (i.e., a document ID) in the corpus that represents the topic of the document. During training, models learn to predict words using one or both of the following models (the authors suggest combining the two methods): • *Distributed memory (DM)* randomly samples consecutive words from a document and predicts a center word using the document ID and context words as inputs. • *Distributed bag of words (DBOW)* takes the document ID as the input and tries to predict randomly sampled words from the document.
Latent Dirichlet Allocation (LDA)	A form of topic modeling. Calculates the probability that each document belongs to one of *t* unobserved (i.e., latent) topics, where *t* is defined *a priori*.

TABLE S3 CV and TF-IDF model performance by board leadership variable

		Vector Type							
		Count Vector				TF-IDF			
Classification Model	**n-gram**	**Train Accuracy**	**Test Accuracy**	**Train/Test Loss**	**AUC**	**Train Accuracy**	**Test Accuracy**	**Train/Test Loss**	**AUC**
Duality									
Logistic regression	2-gram	0.999	0.836	0.163	0.913	0.957	0.831	0.125	0.902
Logistic regression	3-gram	1.000	0.841	0.159	0.915	0.971	0.836	0.135	0.901
Logistic regression	5-gram	1.000	0.854	0.146	0.918	0.988	0.829	0.159	0.894
Logistic regression	10-gram	1.000	0.846	0.154	0.913	0.994	0.816	0.177	0.886
Logistic regression	20-gram	1.000	0.851	0.149	0.912	0.992	0.819	0.173	0.887
Random forest	2-gram	1.000	0.861	0.139	0.918	1.000	0.859	0.141	0.920
Random forest	3-gram	1.000	0.851	0.149	0.917	1.000	0.861	0.139	0.922
Random forest	5-gram	0.986	0.856	0.130	0.915	0.992	0.856	0.136	0.918
Random forest	10-gram	1.000	0.829	0.171	0.909	0.991	0.844	0.147	0.917
Random forest	20-gram	1.000	0.824	0.176	0.901	0.991	0.856	0.135	0.914
Control									
Logistic regression	2-gram	0.994	0.861	0.133	0.904	0.958	0.846	0.113	0.899
Logistic regression	3-gram	0.996	0.861	0.135	0.908	0.969	0.848	0.121	0.905
Logistic regression	5-gram	0.991	0.863	0.128	0.907	0.985	0.848	0.137	0.909
Logistic regression	10-gram	0.999	0.843	0.156	0.907	0.993	0.831	0.162	0.911
Logistic regression	20-gram	0.988	0.836	0.152	0.907	0.968	0.831	0.137	0.905
Random forest	2-gram	1.000	0.846	0.154	0.905	1.000	0.843	0.157	0.901
Random forest	3-gram	1.000	0.831	0.169	0.907	1.000	0.833	0.167	0.905
Random forest	5-gram	1.000	0.833	0.167	0.904	1.000	0.828	0.172	0.903
Random forest	10-gram	1.000	0.823	0.177	0.901	1.000	0.831	0.169	0.899
Random forest	20-gram	1.000	0.828	0.172	0.906	1.000	0.821	0.179	0.891
Collaboration									
Logistic regression	2-gram	0.988	0.833	0.155	0.922	0.949	0.846	0.103	0.922
Logistic regression	3-gram	0.996	0.848	0.147	0.925	0.968	0.838	0.129	0.922
Logistic regression	5-gram	0.998	0.846	0.152	0.929	0.978	0.828	0.150	0.922
Logistic regression	10-gram	0.990	0.858	0.132	0.931	0.984	0.826	0.158	0.919
Logistic regression	20-gram	0.983	0.853	0.129	0.930	0.981	0.828	0.152	0.919
Random forest	2-gram	0.989	0.826	0.163	0.916	0.996	0.833	0.163	0.919
Random forest	3-gram	0.972	0.831	0.141	0.921	0.993	0.833	0.159	0.918
Random forest	5-gram	0.953	0.818	0.135	0.917	0.980	0.833	0.147	0.911
Random forest	10-gram	0.999	0.803	0.195	0.905	0.999	0.836	0.163	0.910
Random forest	20-gram	0.999	0.806	0.193	0.911	0.967	0.821	0.146	0.914

TABLE S4 Word2Vec model performance by board leadership variable

		Vector Size											
		100				200				300			
Classification Model	**n-gram**	**Train Accuracy**	**Test Accuracy**	**Train/Test Loss**	**AUC**	**Train Accuracy**	**Test Accuracy**	**Train/Test Loss**	**AUC**	**Train Accuracy**	**Test Accuracy**	**Train/Test Loss**	**AUC**
Duality													
Logistic regression	2-gram	0.741	0.732	0.009	0.793	0.737	0.730	0.008	0.794	0.735	0.737	-0.002	0.790
Logistic regression	3-gram	0.734	0.732	0.002	0.788	0.736	0.730	0.006	0.796	0.734	0.739	-0.006	0.790
Logistic regression	5-gram	0.752	0.715	0.037	0.799	0.734	0.737	-0.003	0.790	0.733	0.742	-0.009	0.794
Logistic regression	10-gram	0.750	0.725	0.026	0.790	0.737	0.734	0.003	0.793	0.740	0.737	0.003	0.792
Logistic regression	20-gram	0.741	0.737	0.004	0.796	0.738	0.730	0.009	0.789	0.738	0.739	-0.001	0.791
Random forest	2-gram	1.000	0.737	0.263	0.805	0.998	0.744	0.253	0.817	1.000	0.732	0.268	0.809
Random forest	3-gram	1.000	0.757	0.243	0.809	0.999	0.762	0.237	0.808	1.000	0.739	0.261	0.806
Random forest	5-gram	1.000	0.752	0.248	0.816	1.000	0.752	0.248	0.819	1.000	0.759	0.241	0.825
Random forest	10-gram	1.000	0.712	0.288	0.799	1.000	0.747	0.253	0.804	1.000	0.734	0.266	0.812
Random forest	20-gram	1.000	0.742	0.258	0.816	1.000	0.737	0.263	0.815	1.000	0.749	0.251	0.815
Control													
Logistic regression	2-gram	0.768	0.714	0.054	0.764	0.748	0.709	0.039	0.763	0.745	0.701	0.043	0.734
Logistic regression	3-gram	0.755	0.704	0.051	0.759	0.747	0.701	0.045	0.754	0.757	0.709	0.048	0.753
Logistic regression	5-gram	0.762	0.711	0.051	0.756	0.754	0.709	0.045	0.757	0.744	0.694	0.050	0.737
Logistic regression	10-gram	0.769	0.729	0.040	0.769	0.752	0.711	0.041	0.752	0.746	0.699	0.047	0.757
Logistic regression	20-gram	0.757	0.692	0.066	0.765	0.746	0.704	0.042	0.756	0.750	0.704	0.046	0.759
Random forest	2-gram	1.000	0.806	0.194	0.874	1.000	0.794	0.206	0.870	1.000	0.803	0.197	0.865
Random forest	3-gram	1.000	0.799	0.201	0.874	1.000	0.803	0.197	0.863	0.997	0.808	0.188	0.874
Random forest	5-gram	0.996	0.799	0.198	0.871	0.988	0.784	0.205	0.856	0.986	0.786	0.200	0.865
Random forest	10-gram	1.000	0.818	0.182	0.867	1.000	0.806	0.194	0.869	1.000	0.816	0.184	0.861
Random forest	20-gram	1.000	0.799	0.201	0.859	1.000	0.806	0.194	0.867	1.000	0.799	0.201	0.874
Collaboration													
Logistic regression	2-gram	0.778	0.769	0.010	0.838	0.768	0.771	-0.003	0.829	0.769	0.789	-0.020	0.835
Logistic regression	3-gram	0.786	0.769	0.017	0.836	0.776	0.794	-0.018	0.847	0.767	0.776	-0.009	0.836
Logistic regression	5-gram	0.780	0.759	0.022	0.830	0.762	0.771	-0.009	0.838	0.765	0.774	-0.009	0.836
Logistic regression	10-gram	0.773	0.776	-0.003	0.845	0.779	0.784	-0.004	0.831	0.765	0.776	-0.011	0.837
Logistic regression	20-gram	0.782	0.769	0.014	0.840	0.772	0.776	-0.004	0.839	0.767	0.779	-0.011	0.833
Random forest	2-gram	0.998	0.811	0.187	0.896	0.998	0.826	0.172	0.895	0.998	0.826	0.172	0.895
Random forest	3-gram	0.998	0.821	0.177	0.891	0.998	0.826	0.172	0.896	0.998	0.826	0.172	0.897
Random forest	5-gram	0.998	0.823	0.174	0.890	0.998	0.826	0.172	0.895	0.998	0.811	0.187	0.898
Random forest	10-gram	0.998	0.821	0.177	0.895	0.998	0.826	0.172	0.895	0.998	0.826	0.172	0.898
Random forest	20-gram	0.998	0.831	0.167	0.891	0.998	0.831	0.167	0.894	0.998	0.818	0.179	0.893

TABLE S5 Doc2Vec model performance by board leadership variable

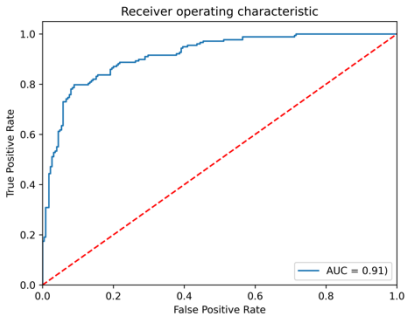
		Vector Size											
		100				200				300			
Classification Model	**n-gram**	**Train Accuracy**	**Test Accuracy**	**Train/Test Loss**	**AUC**	**Train Accuracy**	**Test Accuracy**	**Train/Test Loss**	**AUC**	**Train Accuracy**	**Test Accuracy**	**Train/Test Loss**	**AUC**
Duality													
Logistic regression	2-gram	0.842	0.754	0.088	0.806	0.855	0.749	0.105	0.800	0.845	0.720	0.125	0.778
Logistic regression	3-gram	0.844	0.742	0.102	0.808	0.858	0.749	0.109	0.801	0.862	0.762	0.100	0.808
Logistic regression	5-gram	0.842	0.774	0.068	0.848	0.853	0.752	0.101	0.801	0.850	0.742	0.108	0.812
Logistic regression	10-gram	0.843	0.749	0.094	0.818	0.830	0.737	0.093	0.797	0.895	0.725	0.171	0.781
Logistic regression	20-gram	0.841	0.754	0.087	0.802	0.857	0.757	0.100	0.806	0.852	0.742	0.110	0.812
Random forest	2-gram	1.000	0.749	0.251	0.844	0.999	0.759	0.240	0.831	1.000	0.757	0.243	0.845
Random forest	3-gram	1.000	0.782	0.218	0.867	1.000	0.769	0.231	0.842	1.000	0.757	0.243	0.840
Random forest	5-gram	1.000	0.757	0.243	0.845	1.000	0.764	0.236	0.839	1.000	0.744	0.256	0.833
Random forest	10-gram	1.000	0.742	0.258	0.834	1.000	0.759	0.241	0.849	1.000	0.762	0.238	0.846
Random forest	20-gram	1.000	0.769	0.231	0.843	1.000	0.720	0.280	0.813	1.000	0.747	0.253	0.838
Control													
Logistic regression	2-gram	0.853	0.791	0.062	0.821	0.887	0.791	0.096	0.827	0.883	0.791	0.092	0.824
Logistic regression	3-gram	0.866	0.789	0.078	0.830	0.880	0.774	0.106	0.825	0.886	0.769	0.117	0.817
Logistic regression	5-gram	0.856	0.774	0.083	0.832	0.879	0.776	0.103	0.824	0.879	0.784	0.096	0.817
Logistic regression	10-gram	0.861	0.774	0.087	0.838	0.920	0.746	0.174	0.793	0.890	0.769	0.122	0.816
Logistic regression	20-gram	0.859	0.764	0.095	0.815	0.881	0.766	0.115	0.819	0.886	0.784	0.103	0.811
Random forest	2-gram	1.000	0.808	0.192	0.872	1.000	0.808	0.192	0.866	1.000	0.801	0.199	0.864
Random forest	3-gram	1.000	0.826	0.174	0.881	1.000	0.813	0.187	0.853	1.000	0.816	0.184	0.865
Random forest	5-gram	1.000	0.813	0.187	0.870	1.000	0.818	0.182	0.864	1.000	0.791	0.209	0.855
Random forest	10-gram	1.000	0.803	0.197	0.874	1.000	0.811	0.189	0.859	1.000	0.818	0.182	0.872
Random forest	20-gram	1.000	0.806	0.194	0.880	1.000	0.796	0.204	0.864	1.000	0.796	0.204	0.856
Collaboration													
Logistic regression	2-gram	0.850	0.776	0.074	0.866	0.872	0.766	0.106	0.839	0.873	0.771	0.102	0.833
Logistic regression	3-gram	0.854	0.764	0.091	0.849	0.874	0.766	0.108	0.830	0.866	0.774	0.093	0.844
Logistic regression	5-gram	0.853	0.771	0.081	0.863	0.875	0.759	0.116	0.824	0.871	0.784	0.087	0.834
Logistic regression	10-gram	0.854	0.784	0.071	0.867	0.875	0.769	0.106	0.821	0.877	0.774	0.103	0.833
Logistic regression	20-gram	0.864	0.786	0.078	0.869	0.863	0.764	0.099	0.836	0.879	0.766	0.113	0.833
Random forest	2-gram	1.000	0.808	0.192	0.896	1.000	0.801	0.199	0.885	1.000	0.789	0.211	0.886
Random forest	3-gram	1.000	0.794	0.206	0.894	0.999	0.786	0.213	0.886	1.000	0.828	0.172	0.887
Random forest	5-gram	1.000	0.801	0.199	0.896	1.000	0.786	0.214	0.884	1.000	0.816	0.184	0.885
Random forest	10-gram	0.998	0.813	0.184	0.903	1.000	0.789	0.211	0.878	1.000	0.818	0.182	0.897
Random forest	20-gram	1.000	0.796	0.204	0.894	0.999	0.801	0.198	0.894	1.000	0.799	0.201	0.883

TABLE S6 LDA model performance by board leadership variable

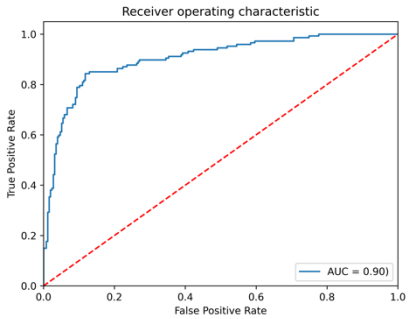
		Topic Number											
		100				200				300			
Classification Model	**n-gram**	**Train Accuracy**	**Test Accuracy**	**Train/Test Loss**	**AUC**	**Train Accuracy**	**Test Accuracy**	**Train/Test Loss**	**AUC**	**Train Accuracy**	**Test Accuracy**	**Train/Test Loss**	**AUC**
Duality													
Logistic regression	2-gram	0.690	0.700	-0.009	0.769	0.741	0.752	-0.011	0.820	0.708	0.715	-0.006	0.808
Logistic regression	3-gram	0.673	0.672	0.000	0.735	0.705	0.692	0.013	0.776	0.725	0.715	0.010	0.817
Logistic regression	5-gram	0.628	0.643	-0.015	0.718	0.701	0.695	0.007	0.774	0.682	0.705	-0.023	0.752
Logistic regression	10-gram	0.619	0.677	-0.058	0.717	0.687	0.692	-0.005	0.774	0.666	0.712	-0.046	0.789
Logistic regression	20-gram	0.657	0.658	0.000	0.733	0.655	0.643	0.012	0.740	0.709	0.707	0.002	0.763
Random forest	2-gram	0.911	0.680	0.231	0.753	1.000	0.715	0.285	0.779	0.836	0.695	0.141	0.761
Random forest	3-gram	0.865	0.635	0.230	0.711	0.879	0.667	0.211	0.757	0.999	0.705	0.295	0.775
Random forest	5-gram	0.821	0.640	0.181	0.698	0.827	0.705	0.123	0.756	0.998	0.677	0.320	0.743
Random forest	10-gram	0.825	0.630	0.195	0.708	0.839	0.663	0.177	0.715	0.950	0.710	0.240	0.748
Random forest	20-gram	0.837	0.650	0.187	0.712	0.811	0.648	0.164	0.716	1.000	0.655	0.345	0.720
Control													
Logistic regression	2-gram	0.726	0.672	0.054	0.732	0.730	0.682	0.048	0.752	0.783	0.749	0.035	0.813
Logistic regression	3-gram	0.718	0.692	0.027	0.731	0.749	0.692	0.058	0.766	0.774	0.716	0.058	0.756
Logistic regression	5-gram	0.666	0.634	0.032	0.633	0.736	0.684	0.052	0.749	0.774	0.739	0.035	0.772
Logistic regression	10-gram	0.705	0.674	0.031	0.723	0.745	0.664	0.081	0.717	0.754	0.692	0.063	0.734
Logistic regression	20-gram	0.704	0.662	0.043	0.708	0.729	0.699	0.030	0.737	0.739	0.689	0.050	0.750
Random forest	2-gram	0.999	0.739	0.260	0.801	1.000	0.744	0.256	0.808	1.000	0.774	0.226	0.830
Random forest	3-gram	0.998	0.761	0.237	0.803	0.999	0.754	0.246	0.798	0.999	0.751	0.248	0.830
Random forest	5-gram	0.991	0.744	0.248	0.753	0.992	0.756	0.236	0.794	0.999	0.731	0.267	0.799
Random forest	10-gram	0.999	0.736	0.263	0.759	0.999	0.714	0.285	0.770	0.993	0.739	0.254	0.788
Random forest	20-gram	0.999	0.761	0.238	0.798	0.997	0.761	0.236	0.812	0.992	0.816	0.176	0.842
Collaboration													
Logistic regression	2-gram	0.717	0.689	0.028	0.759	0.733	0.677	0.056	0.757	0.763	0.719	0.044	0.787
Logistic regression	3-gram	0.713	0.692	0.021	0.768	0.729	0.697	0.032	0.794	0.736	0.701	0.034	0.803
Logistic regression	5-gram	0.689	0.669	0.020	0.759	0.735	0.692	0.043	0.764	0.767	0.714	0.053	0.798
Logistic regression	10-gram	0.654	0.654	0.000	0.734	0.715	0.709	0.006	0.761	0.733	0.709	0.024	0.796
Logistic regression	20-gram	0.647	0.570	0.078	0.649	0.693	0.667	0.026	0.719	0.754	0.704	0.050	0.799
Random forest	2-gram	0.998	0.746	0.251	0.843	0.999	0.754	0.245	0.832	0.999	0.746	0.252	0.840
Random forest	3-gram	0.825	0.684	0.141	0.770	0.998	0.726	0.271	0.815	0.998	0.789	0.210	0.865
Random forest	5-gram	0.994	0.687	0.307	0.770	0.997	0.711	0.285	0.796	0.999	0.701	0.297	0.813
Random forest	10-gram	0.994	0.682	0.313	0.760	0.983	0.704	0.279	0.788	0.998	0.726	0.272	0.825
Random forest	20-gram	0.999	0.692	0.307	0.766	0.999	0.679	0.320	0.789	0.998	0.731	0.267	0.823

TABLE S7 ROC curves and confusion matrices for final selected models

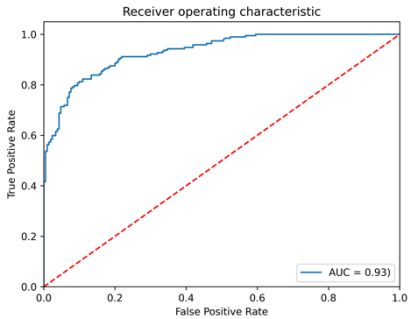
ROC curve



AUC = 0.915



AUC = 0.904



AUC = 0.931

Confusion Matrix

		CEO duality		Control		Collaboration	
		Predicted class		Predicted class		Predicted class	
		0	1	0	1	0	1
True class	0	206	19	229	26	194	16
	1	39	139	30	117	41	151

TABLE S8 Confusion matrices for predicted scores and post-hoc validation

		CEO duality		Control		Collaboration	
Confusion Matrix		Predicted score		Predicted score		Predicted score	
		0	1	0	1	0	1
True class	0	96	10	41	16	39	12
	1	17	77	3	45	2	52
Overall Agreement		87%		82%		87%	
Cohen's Kappa (p-value)		0.73 (.000)		0.64 (.000)		0.73 (.000)	

TABLE S9 Database summary

Variable	Description	Type	N	Min value	Max value
cik	CIK number	long	22,388		
year	Fiscal year of DEF14a	double	22,388		
filingmonth	Month company filed DEF14a	byte	22,388		
filingday	Day company filed DEF14a	byte	22,388		
textid	Sample-specific unique identifier for passage from DEF14a	long	22,388		
text	Full text of passage from DEF14a	string	22,388		
duality_prob	Predicted probability CEO duality = 1	float	22,388	0	1
duality_bin	Binary value of CEO duality	byte	22,388	0	1
control_prob	Predicted probability control orientation = 1	float	13,900	0	1
control_bin	Binary value of control orientation	byte	13,900	0	1
collab_prob	Predicted probability collaboration orientation = 1	float	13,900	0	1
collab_bin	Binary value for collaboration orientation	byte	13,900	0	1

TABLE S10 Variables used in the research application[a]

Variable	Source	Description
Study Variables		
CEO dismissal	Gentry et al (2021) dataset	Coded 1 for all departures that Gentry et al. (2021) classified as involuntary (i.e., departure codes "3" and "4"), and "0" otherwise.
Return on Assets (ROA)	Compustat	Net income divided by total assets (Gentry et al., 2021; Shen and Cannella, 2002)
CEO duality		Coded 1 if the CEO is also board chair, 0 if not.
Collaboration orientation[b]	Current dataset	Coded 1 if the collaboration orientation was identified in the proxy statement text, 0 if not.
Control orientation[b]	Current dataset	Coded 1 if the control orientation was identified in the proxy statement text, 0 if not.
Controls		
Total stock returns	Compustat	Annual share price appreciation plus dividends, divided by beginning price (Gentry et al., 2021; Wiersema et al., 2011).
Average analyst recommendation	IBES	Performance-adjusted average analyst stock recommendations for the fiscal year (reverse coded so that a higher score suggests higher analyst recommendation) (Wiersema et al., 2011).
Leverage	Compustat	Total debt divided by total shareholders' equity.
Firm size	Compustat	Log-transformed firm revenues.
Ln(R&D expenses)	Compustat	Log-transformed R&D expenses (zero-imputed when missing).
CEO age	Execucomp	Age of the CEO in number of years.
CEO tenure	Execucomp	Number of years the CEO has held that position within the firm.
CEO is female	Execucomp	Coded 1 if the CEO is female and 0 otherwise.
Board size	ISS	Total number of directors on the board.
Outside directors	ISS	Count of outside (independent) directors on the board.
Female directors	ISS	Count of female (independent) directors on the board.
Industry fixed effects	Compustat	Industry dummy variables, based on 2-digit SIC codes.
Year fixed effects	All	Year dummy variables.

[a] Except for industry and firm fixed effects, all independent and control variables are measured in t - 1

[b] Collaboration and control orientations are only measured when CEO and board chair positions are separated.

TABLE S11 Descriptive statistics and pairwise correlations for research application[a]

	Variable	N	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	CEO dismissal	6,967	0.03	0.17															
2	Duality	6,967	0.45	0.50	-0.03														
3	Control orientation[b]	3,733	0.39	0.49	-0.03	-													
4	Collaboration orientation[b]	3,733	0.44	0.50	0.02	-	0.10												
5	ROA	6,967	0.05	0.09	-0.06	0.03	0.04	-0.04											
6	Total stock returns	6,967	0.87	23.64	-0.01	-0.03	0.01	-0.02	0.00										
7	Average analyst recommendation	6,967	-0.05	0.45	-0.01	0.02	0.05	0.01	0.07	0.04									
8	Leverage	6,967	1.10	28.20	0.00	-0.01	-0.01	-0.01	-0.02	0.00	0.00								
9	Firm size	6,967	7.73	1.56	0.02	0.13	-0.11	-0.08	0.11	0.08	0.04	0.00							
10	Ln(R&D expenses)	6,967	2.05	2.51	0.02	0.02	0.03	0.05	0.09	-0.02	0.15	0.01	0.17						
11	CEO age	6,967	56.83	7.08	0.00	0.20	0.00	-0.09	0.01	0.11	-0.03	-0.01	0.08	-0.08					
12	CEO tenure	6,967	7.82	7.38	-0.03	0.23	-0.01	-0.05	0.03	0.14	0.00	-0.01	-0.10	-0.06	0.43				
13	CEO is female	6,967	0.04	0.19	0.01	-0.03	0.01	0.03	0.02	-0.01	-0.04	0.00	0.02	-0.01	-0.05	-0.09			
14	Board size	6,967	9.17	2.13	0.01	0.08	-0.07	-0.06	0.04	0.04	-0.01	0.00	0.58	0.11	0.07	-0.11	0.00		
15	Inside directors	6,967	1.82	1.02	-0.03	-0.07	-0.14	-0.03	0.04	0.06	-0.05	-0.02	0.04	-0.14	0.07	0.08	-0.02	0.27	
16	Female directors	6,967	1.40	1.10	0.02	0.08	-0.04	-0.07	0.05	0.03	-0.04	0.01	0.47	0.10	0.02	-0.14	0.23	0.54	-0.03

[a] Confidence intervals for pairwise correlations are available from the authors by request.
[b] Control and collaboration orientations are conditional on the value of CEO duality.

TABLE S12 Marginal effect analysis for research application

	Effect	**SE**	***z***	***p***
CEO duality				
No	-0.067	0.023	-2.95	0.003
Yes	-0.078	0.022	-3.48	0.001
Control orientation				
No	-0.042	0.028	-1.49	0.135
Yes	-0.102	0.040	-2.56	0.011
Collaboration orientation				
No	-0.081	0.027	-3.00	0.003
Yes	-0.046	0.036	-1.30	0.194