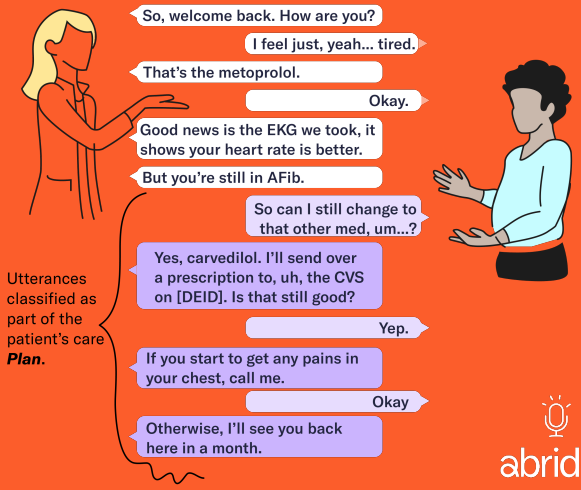
Towards Fairness in Classifying Medical Conversations into SOAP Sections



Benefits of classifier:

Surfacing utterances classified as ***Plan*** improves recall and understanding of patient care plans.

Are there disparities in classifier performance? (and in allocation of benefits)

Metrics:

Average Odds Difference (AOD): $\frac{(FPR_{disadv}-FPR_{adv})+(TPR_{disadv}-TPR_{adv})}{2}$

Equal Opportunity Ratio (EOR): $\frac{TPR_{disadv}}{TPR_{adv}}$

False Omission Rate Ratio (FORR): $\frac{FOR_{disadv}}{FOR_{adv}}$

Protected Attributes and Groups:

Protected Attribute	Disadvantaged Group	Advantaged Group
Race/Ethnicity	Black Hispanic Asian	White White White
Gender	Female (Patient) Female (Physician)	Male (Patient) Male (Physician)
Race/Ethnicity + Gender	Black female Hispanic female	White male White male
Socioeconomic	Unemployed Retired Nursing home Incarcerated Medicaid Uninsured	Full-time job Full-time job Living at home Living at home Private insurance Private insurance
Obesity	>=250 lbs.	<250 lbs.
Mental health	Psychiatrist (Physician)	Other specialty (Physician)
Location	Other U.S. state	FL, CA, and NY

Analysis

- **Language:** By measuring association between n-grams and labels (local mutual information), we find lexical cues are different for the groups with disparate FORR, suggesting a different distribution of medical providers.
- **Medical provider:** We recalculate metrics after omitting visits to certain medical providers that are slightly more frequent in the disadvantaged groups.

Results

- Group disparities identified for 7 (out of 90) cases for one metric (FORR).
- Of these, 3 are in the *Plan* class (which provides benefits to users).
- Omitting indicated medical provider eliminates disparity (hatched green bars)

Protected Attribute	Disadvantaged Group	Advantaged Group	Omitted Medical Provider	1+AOD	EOR	FORR
Race/Ethnicity	Asian	White	Clinical cardiologist, Ophthalmogolist	1.00	.97	.83
Race/Ethnicity + Gender	Hispanic Female	White Male	Allergist	.99	.95	.82
Socioeconomic	Incarcerated	Living at home	Infect. disease specialist	.95	.95	.82

Conclusion

Disparities in classifier performance can be traced back to different types of medical visits. This finding highlights the importance of understanding the differences already present in datasets and how these can affect a model's ability to allocate equal benefit.