

NLI: Identifying the CHARACTERistics of a Native Language

Introduction

We show that it is possible to learn to identify the mother tongue of TOEFL test takers from the content of the essays they wrote. Our method uses a character level model based on convolutional neural networks to predict the most probable native language from a set of 11 possibilities.

Background

- Initial thoughts
 - How would a human approach this task?
 - Misspellings
 - Specific to each language
 - Words? Characters?
- Character embedding motivation
 - NLI Shared Task
 - Groningen (3rd place, 1-9 character n-grams)
 - Google Brain 2016
 - Evaluated on 1B Word benchmark
 - Outperformed state of the art word embedding technologies
 - Capture complex aspects of language

Background

■ Initial Thoughts

How would a human approach this task?

Misspellings: Specific to each language? Words? Characters?

■ Character Embedding Motivation

NLI Shared Task: Groningen (3rd place, 1-9 character n-grams)

Google Brain 2016:

- Evaluated on 1B Word benchmark

- Outperformed state of the art word embedding technologies

- Capture complex aspects of language

Model

Model Overview

■ A Document is Considered a Sequence of Sentences

■ A Sentence is Considered a Sequence of Characters

■ Learn Vector Repr. of Each Unit

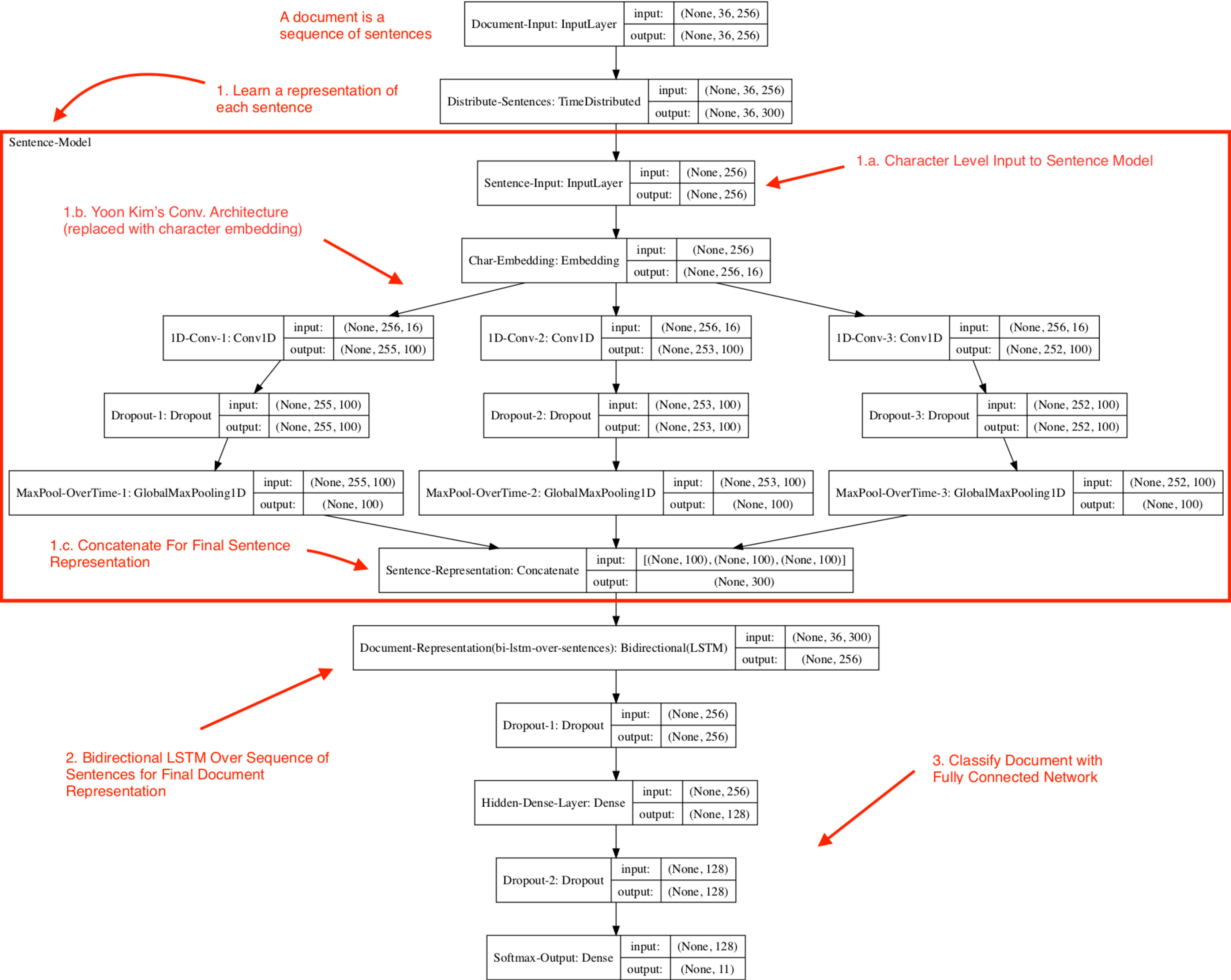
Character (Initial Embedding Layer)

Sentence (Variant of Yoon Kim’s Conv. Architecture)

Document (Bidirectional LSTM)

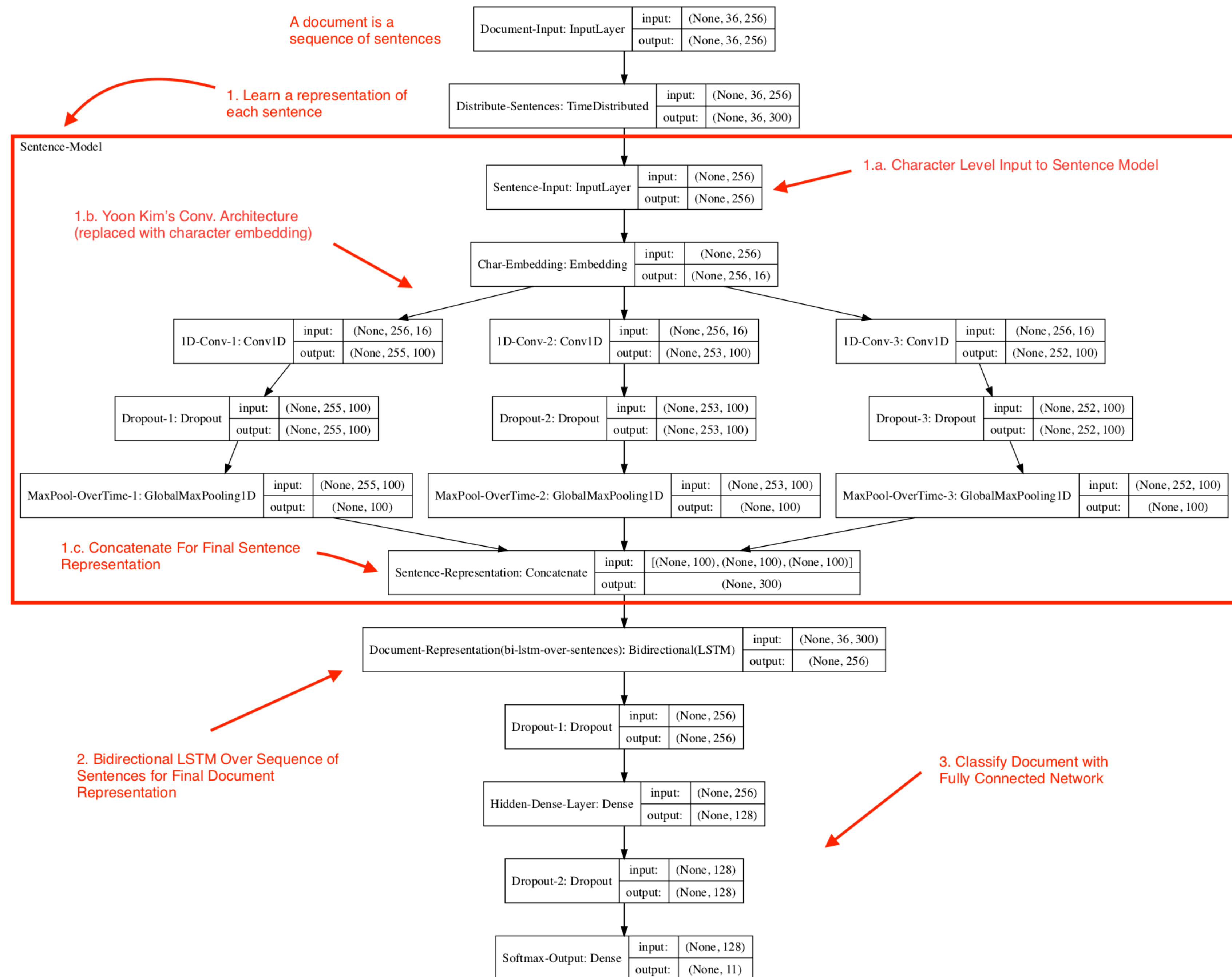
■ Classify Based on Final Doc Repr.

■ Dense Layers with Softmax Output for Class Probabilities



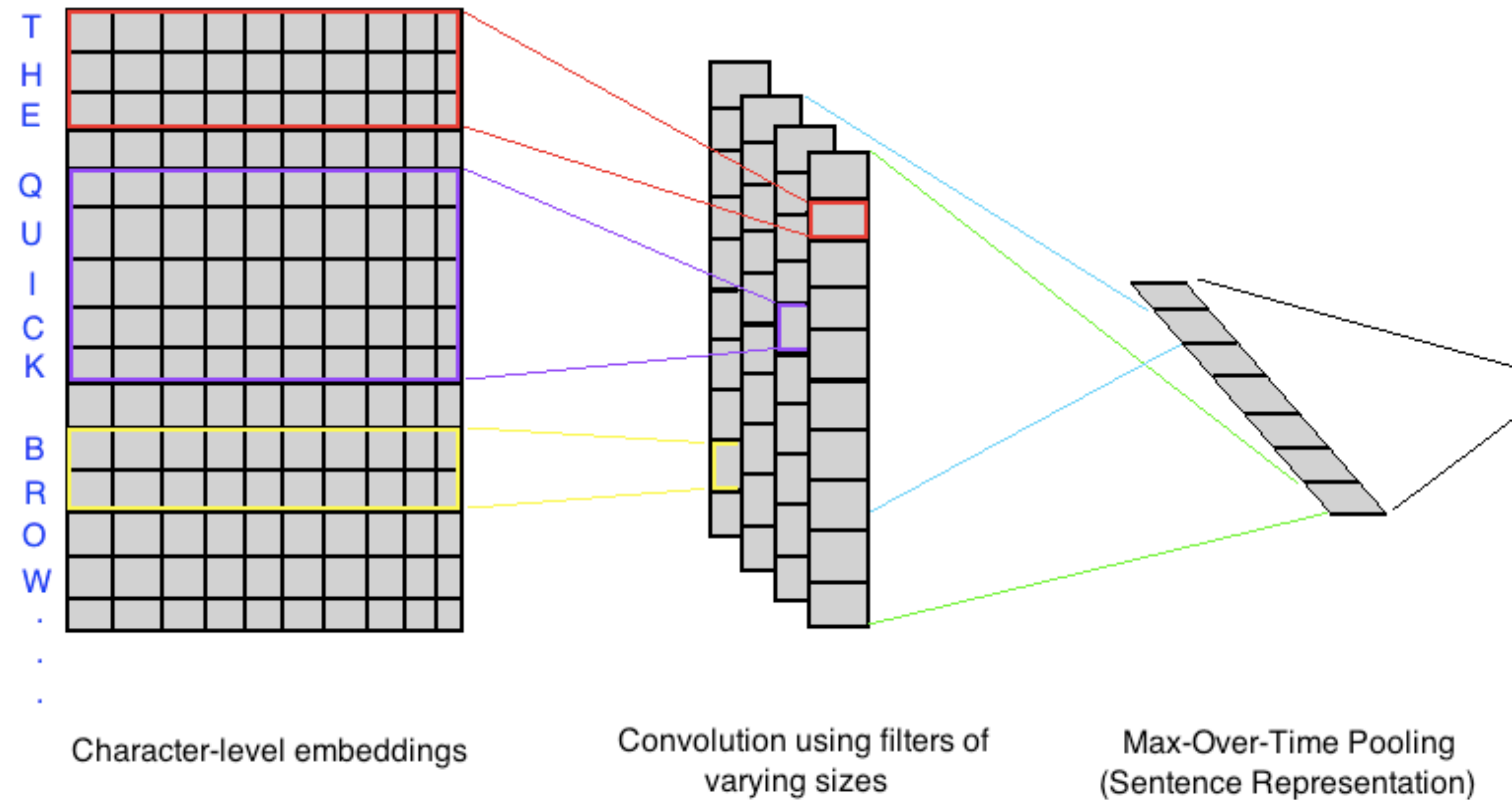
Model Overview

- A document is considered a sequence of sentences
- A sentence is considered a sequence of characters
- Learn vector representations of each unit
 - Character (Initial embedding Layer)
 - Sentence (Variant of Yoon Kim's Conv. Architecture)
 - Document (Bidirectional LSTM)
- Classify based on final document representation
- Dense Layers with Softmax output for class probabilities



Sentence Representation

- Input Learned Character Embedding
- 1D Convolutions by 100 filters of varying sizes (2, 4, 5)
 - Capture short-sequences of characters and small words
 - Identify things like misspellings
- Max-Pooling Over-Time
 - Remember most important aspects
- Concatenate Values for Vector Representation

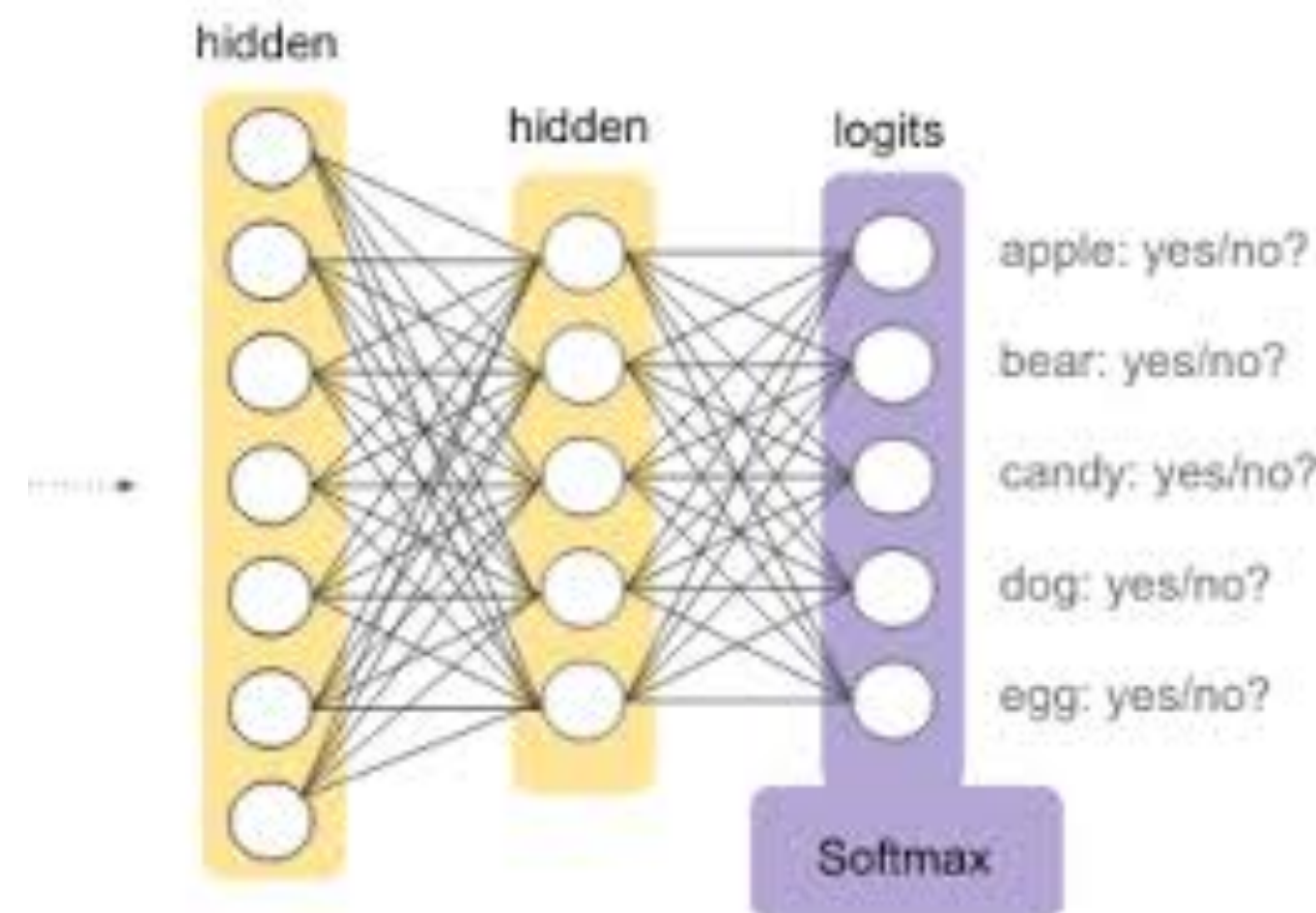
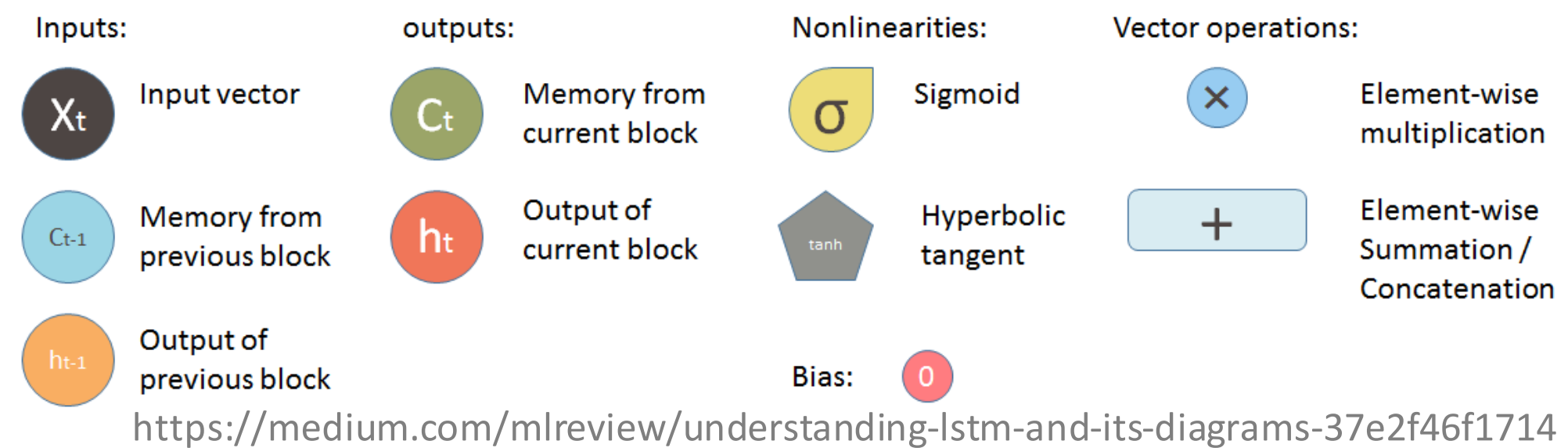
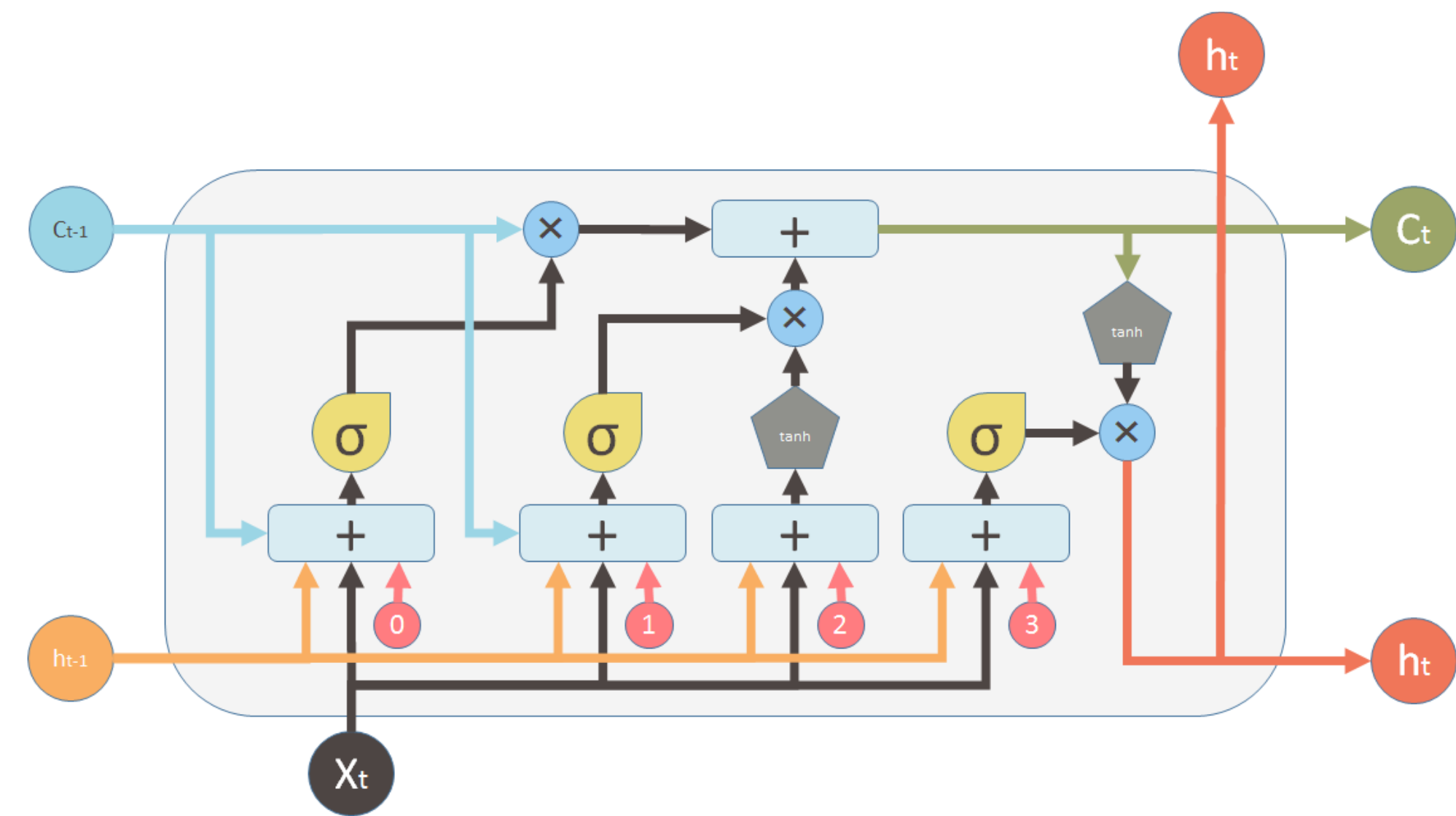


Document Representation

- Input Learned Sentence Repr.
- LSTM over sequence of sentences for document representation
- Each cell receives information from previous cells and current input from sequence

Use of sigmoid layers to decide what old information should be forgotten and new information added

- Dense Layers with Softmax output for class probabilities
- Dropout throughout model for regularization



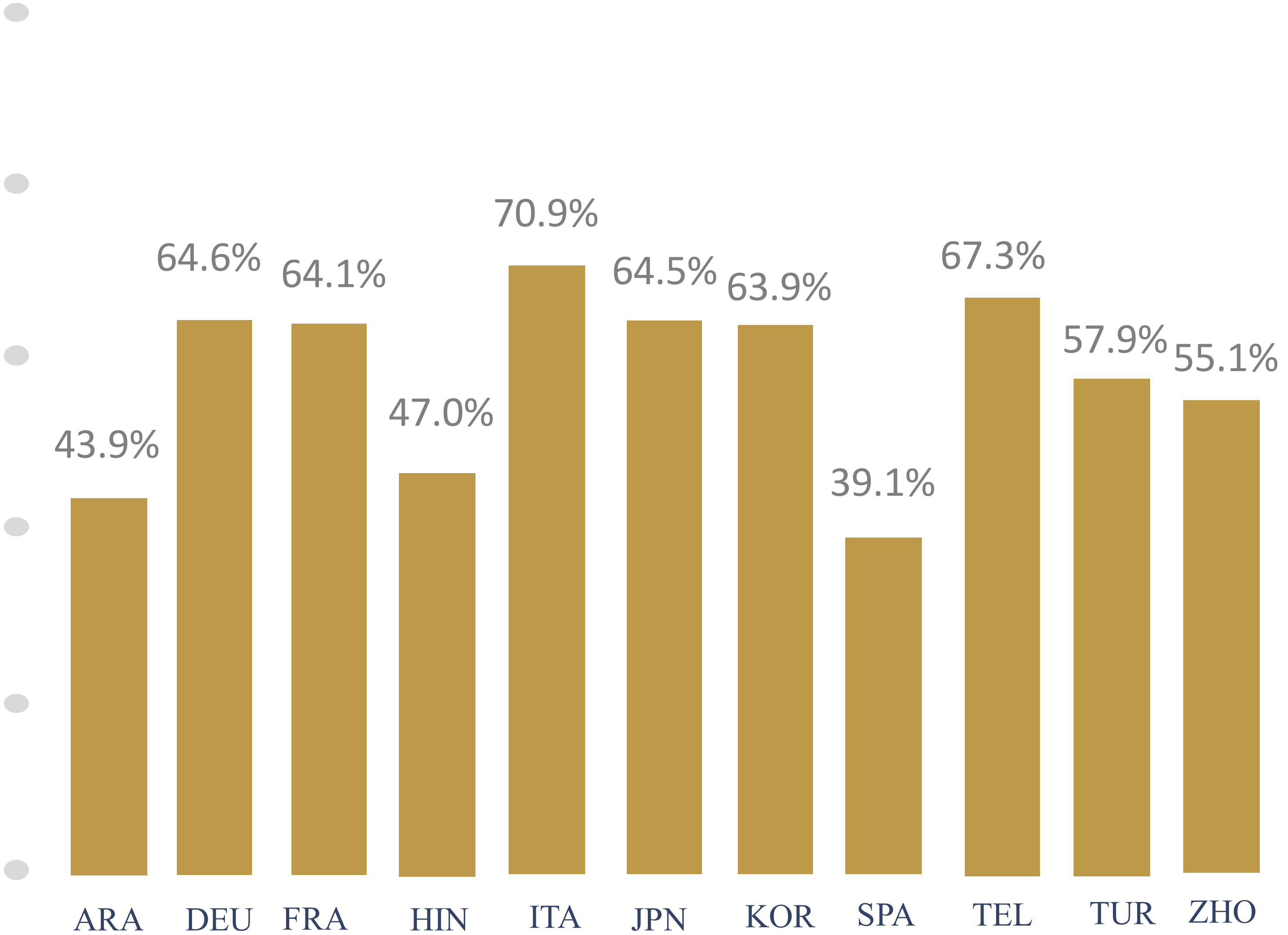
$$\sigma(x_j) = \frac{e^{x_j}}{\sum_i e^{x_i}}$$

<https://jamesmccaffrey.wordpress.com/2016/03/04/the-max-trick-when-computing-softmax/>

<https://developers.google.com/machine-learning/crash-course/multi-class-neural-networks/softmax>

Results

Precision on Test Set



Relatively High Precision on Italian

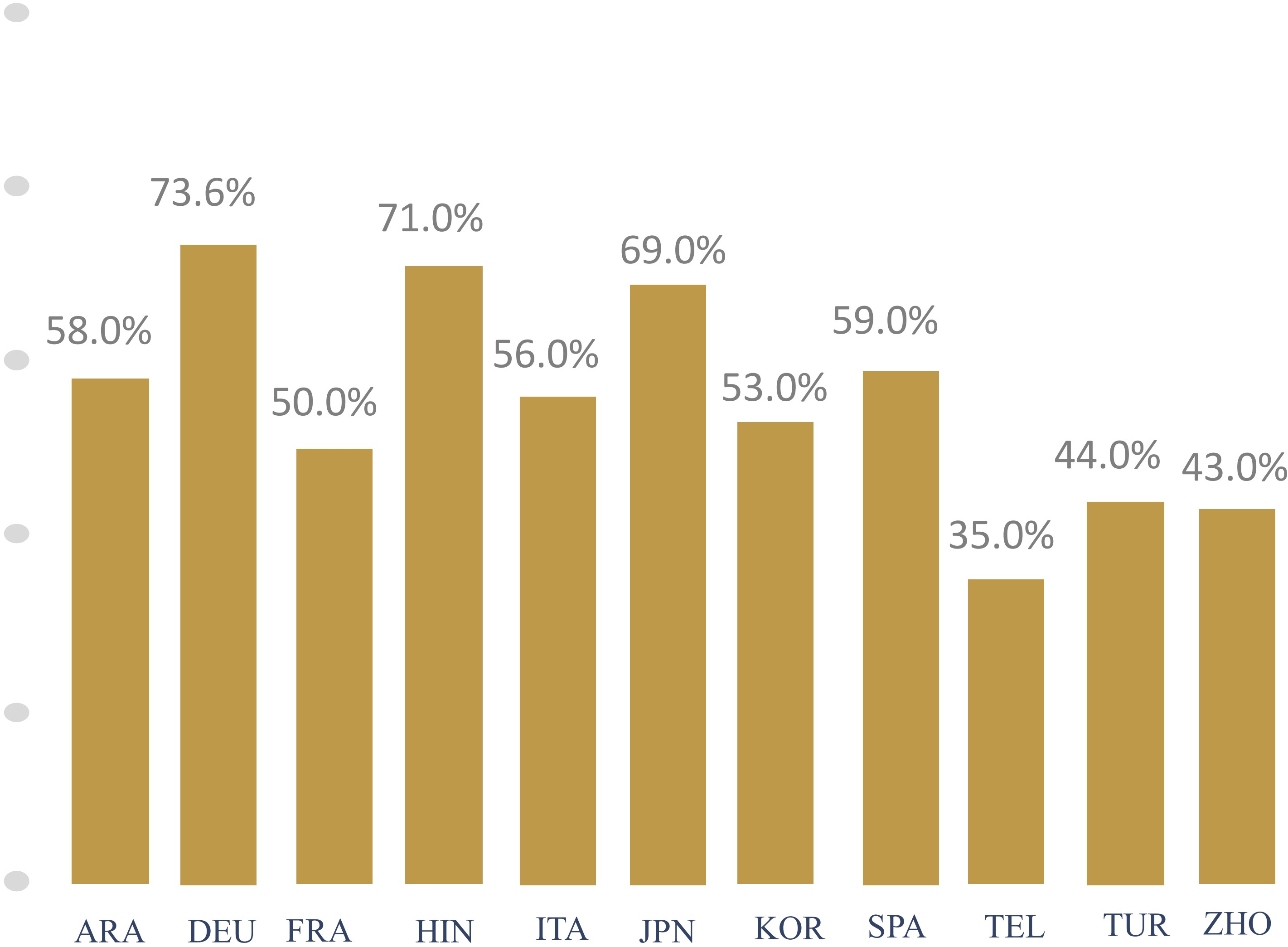
According to phonotactics (“arrangements of sounds”), Italian allows V, CV, CCV, CCCV for a syllable. Every syllable ends with a vowel. That includes the last syllable. (vowel, consonant)

I recently read an article about supporting groups, which have to face the situation of very low activity participating rates among their members and lack of new applications. Since I also lead a supporting group for keeping clean our city park, I can confirme this tendence and I agree with the statement.

I think that in twenty years there will be flying cars !
Today the car is become a status symbol , there is n't a family without a car , but the problem is that there are a lot of people who prefere to have a beautiful car than a sweet home .

easier to remember the facts because they would have a logical order.
For istance, once I tried to remeber all the facts which occurred during the French

Recall on Test Set



German

550/900 High

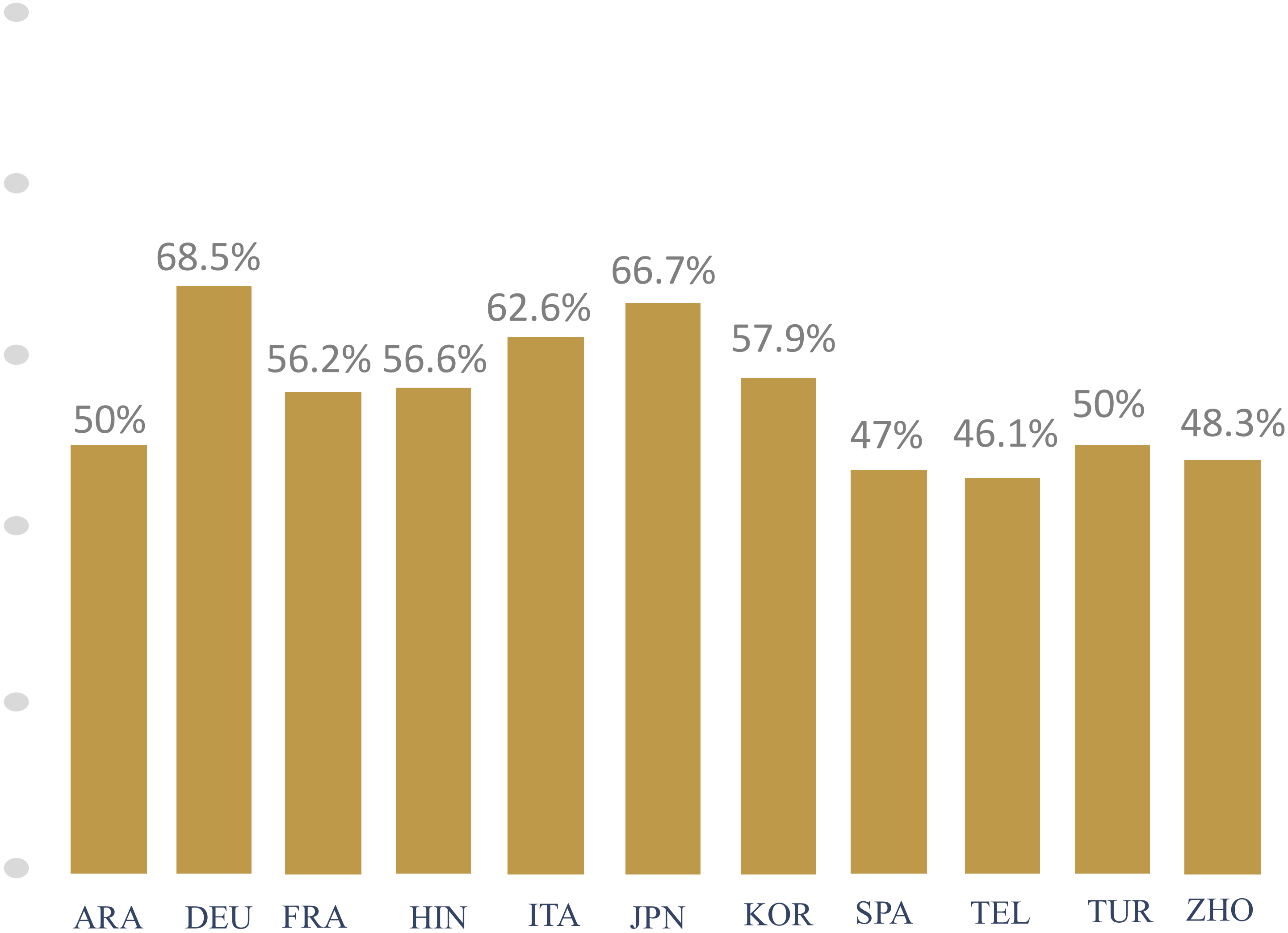
Hindi

526/900 High

Japanese

160/900 High

F1 Score on Test Set



Thanks!

ANY QUESTIONS?

Team Members:

Anthony Sicilia
Brian Falkenstein
Yunkai Tang

Brief Introduction

Here something



SECTION BREAK

Successful people do what
unsuccessful people are not
willing to do.

Hello!

David Porter

Refers to a good or service being offered by a company. Ideally, a product should meet a certain consumer demand, or it should be so compelling that consumers believe they need it.

Transition Headline

Refers to a good or service being offered by a company.

“

Successful people do what unsuccessful people are not willing to do. Don't wish it were easier; wish you were better.

”

This Is A Slide Title

■ The Problem

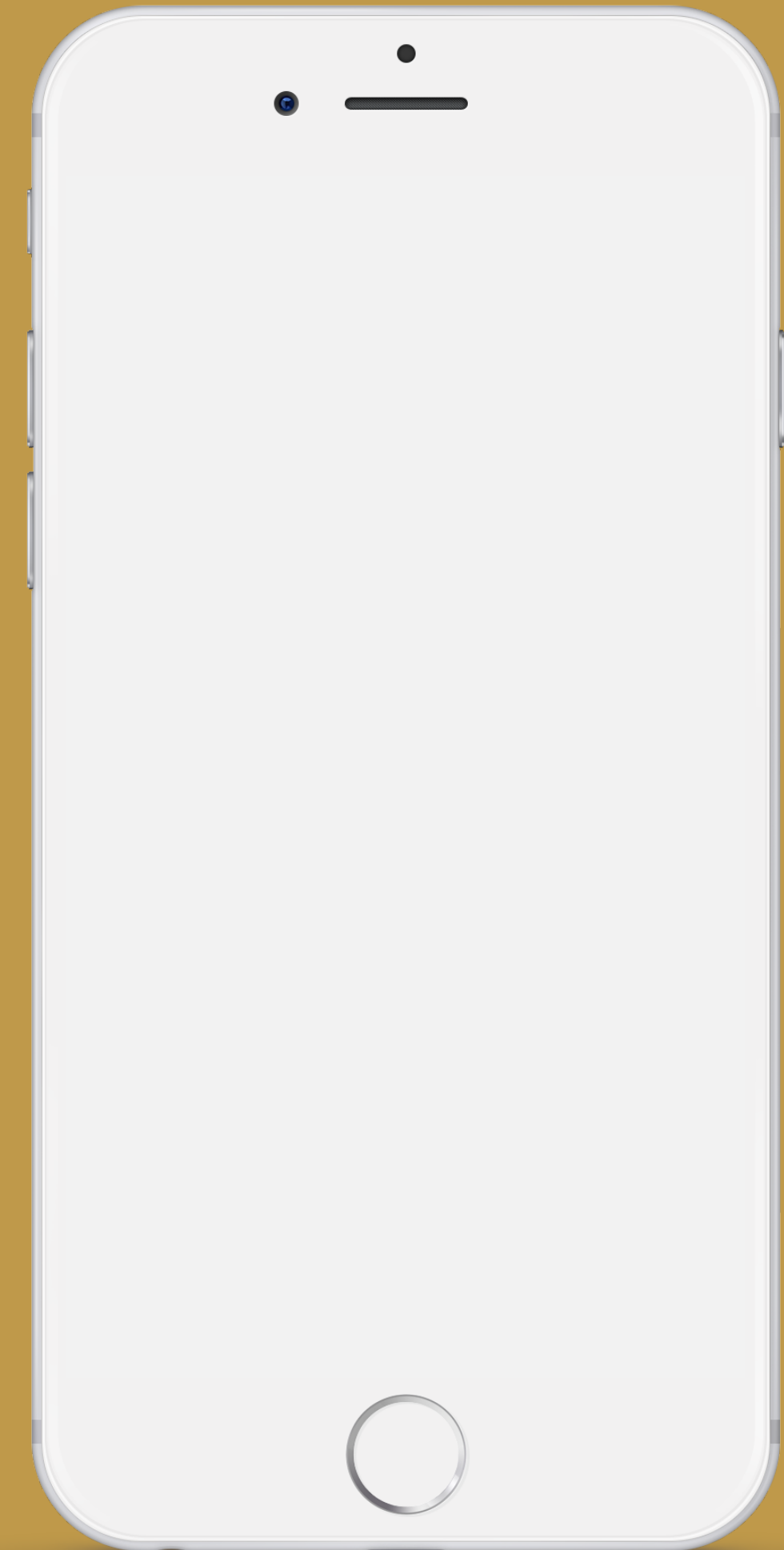
Some fifty years ago there was a curious case of whale trover litigated.
Likewise a fish is technically fast when it bears a waif.

■ The Solution

Some fifty years ago there was a curious case of whale trover litigated.
Likewise a fish is technically fast when it bears a waif.

■ Service

Some fifty years ago there was a curious case of whale trover litigated.
Likewise a fish is technically fast when it bears a waif.

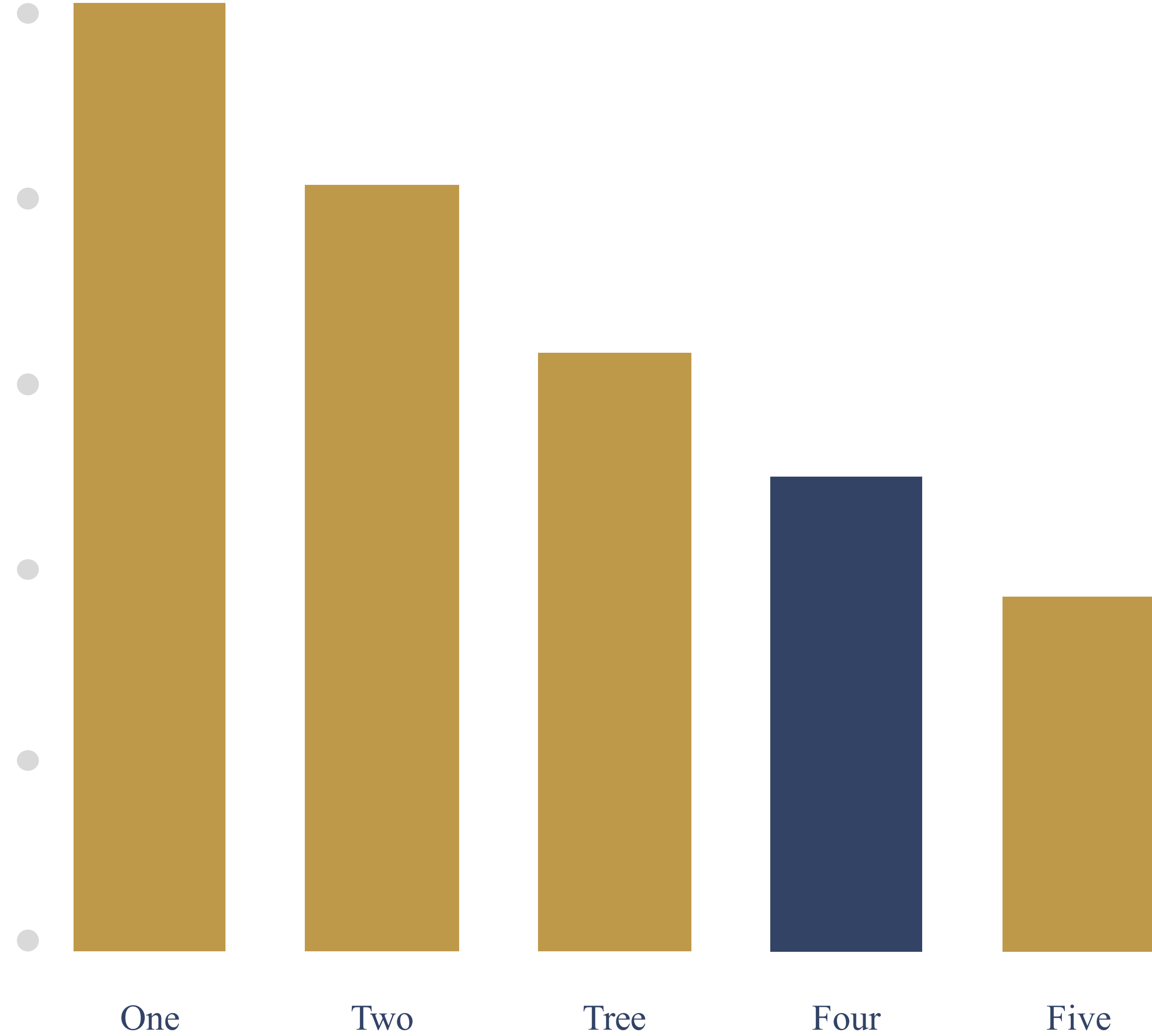




SECTION BREAK

Successful people do what
unsuccessful people are not
willing to do.

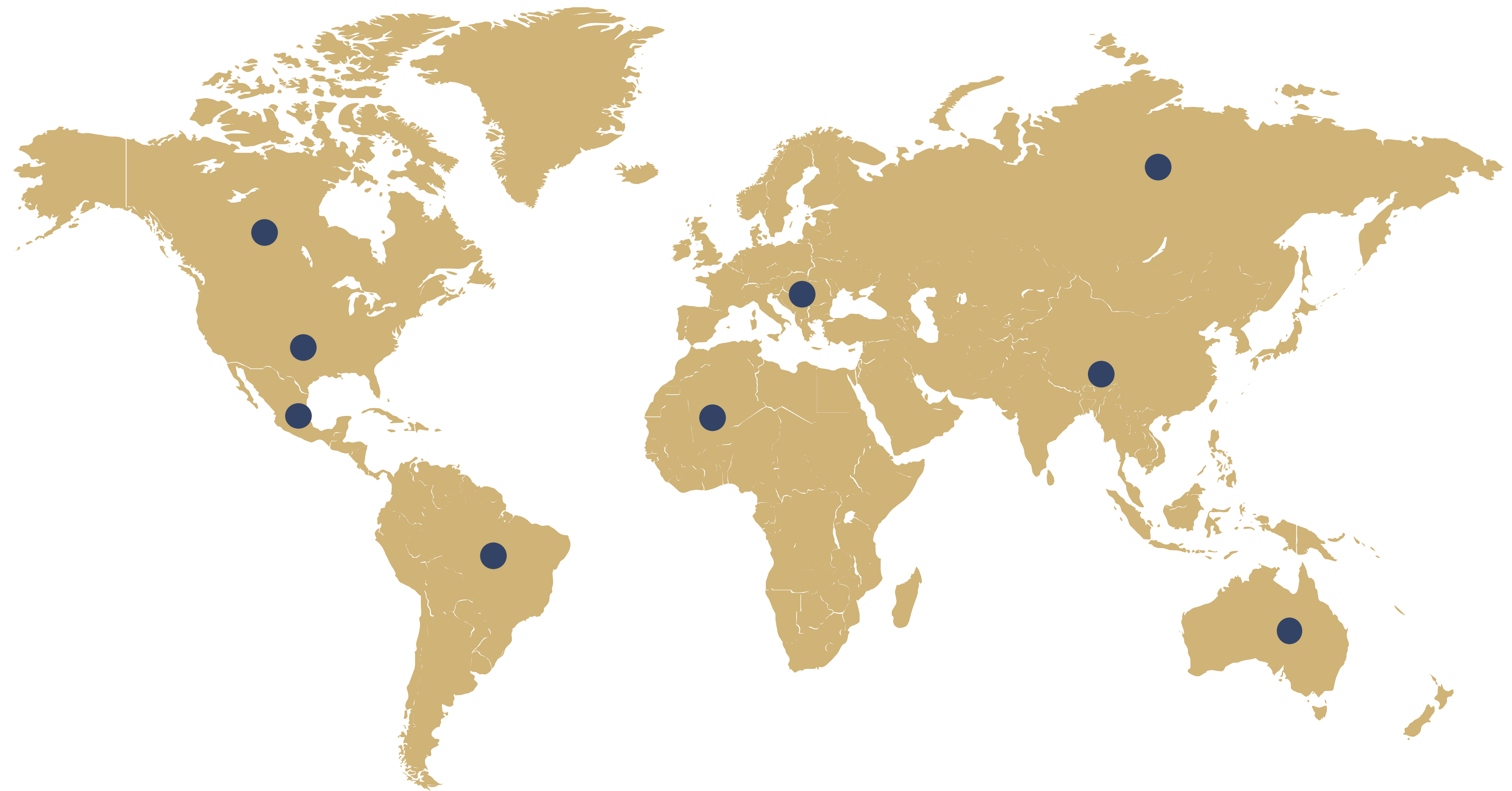
Use Charts To Explain Your Ideas



Use Charts To Explain Your Ideas

One	Two	Tree	Four
\$4390	\$193	\$67	\$290

Maps



2,895,730.00

That's a big number

285,637

That's a big
number

7,845,958

That's a big
number

80%

That's a big
number

Our process is easy

Unsuccessful people are not willing to do. Don't wish it were easier; wish you were better.

One → Two → Tree

Bulleted List Sample

■ The Problem

Some fifty years ago there was a curious case of whale trover litigated. Likewise a fish is technically fast when it bears a waif. These are scientific commentaries; but the commentaries of the whale men themselves sometimes consist in hard words and harder knocks—the Coke-upon-Littleton of the fist.

■ The Solution

Some fifty years ago there was a curious case of whale trover litigated. Likewise a fish is technically fast when it bears a waif. These are scientific commentaries; but the commentaries of the whale men themselves sometimes consist in hard words and harder knocks—the Coke-upon-Littleton of the fist.

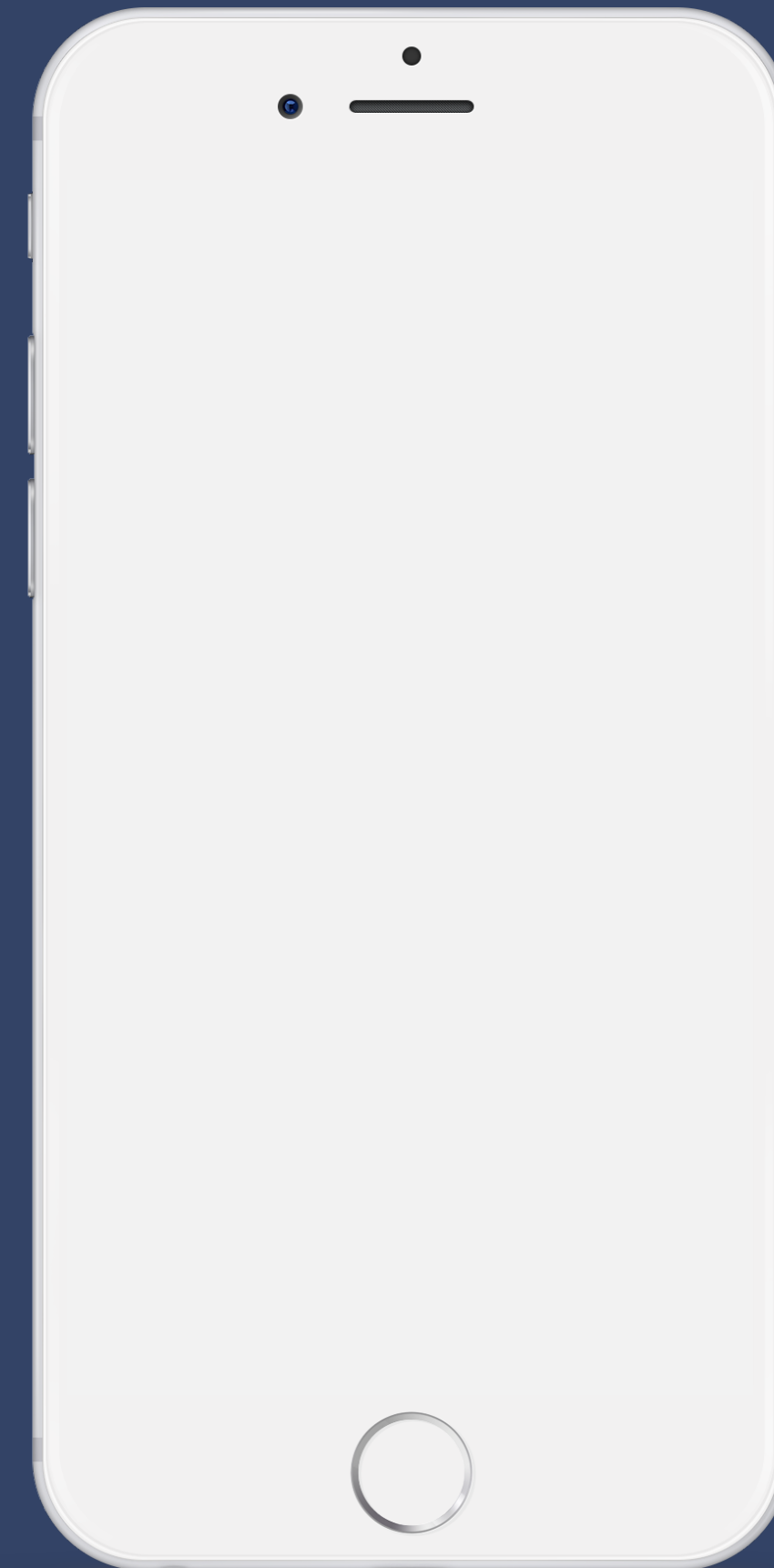
■ Service

Some fifty years ago there was a curious case of whale trover litigated. Likewise a fish is technically fast when it bears a waif. These are scientific commentaries; but the commentaries of the whale men themselves sometimes consist in hard words and harder knocks—the Coke-upon-Littleton of the fist.

iPhone Project

iPhone

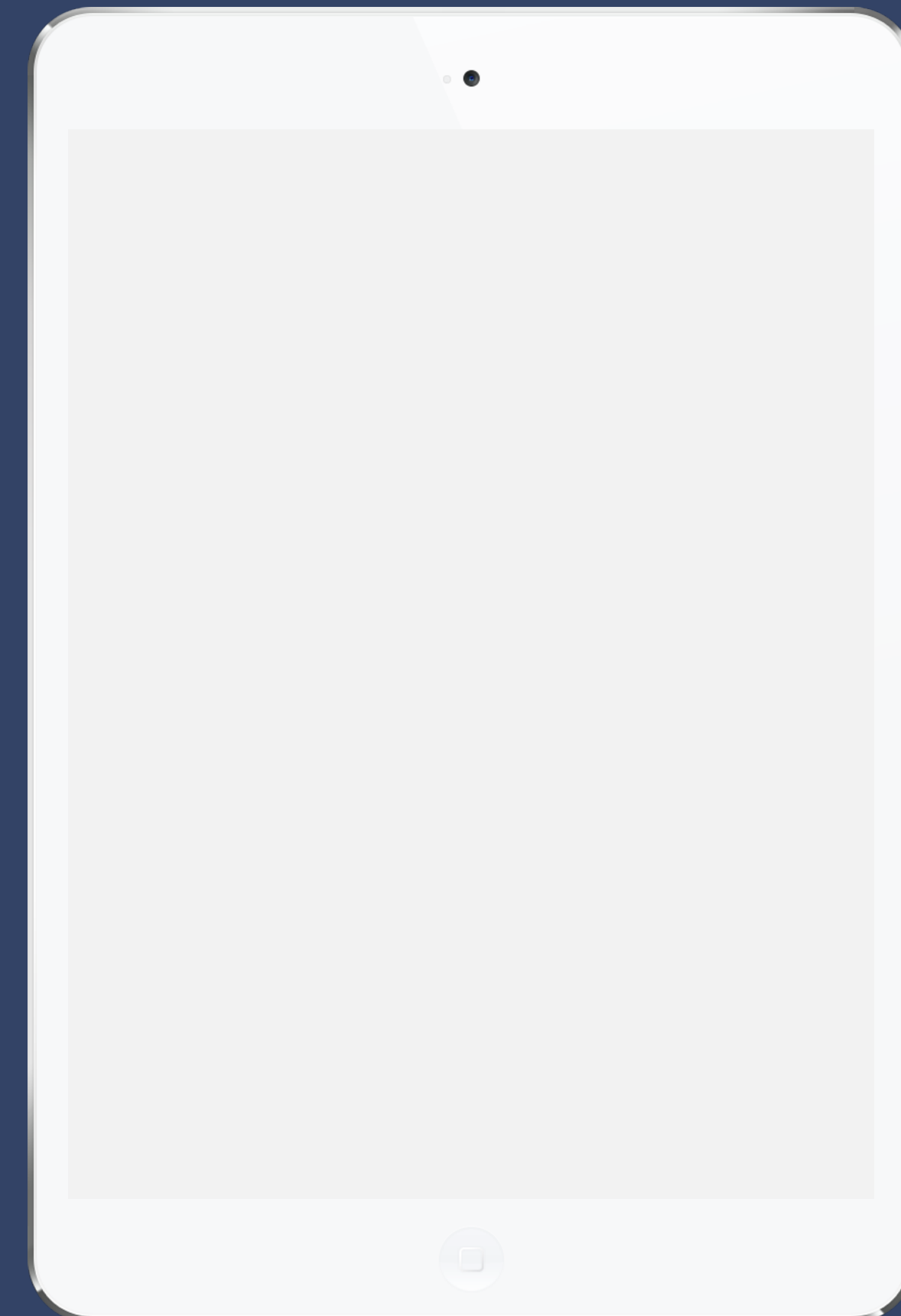
Refers to a good or service being offered by a company. Ideally, a product should meet a certain consumer demand, or it should be so compelling that consumers believe they need it. To be successful, marketers should understand the life cycle of a product, and business executives should have a plan for dealing with products at every stage of their life cycles. Some fifty years ago there was a curious case of whale-trover litigated in England.



iPad Project

iPad

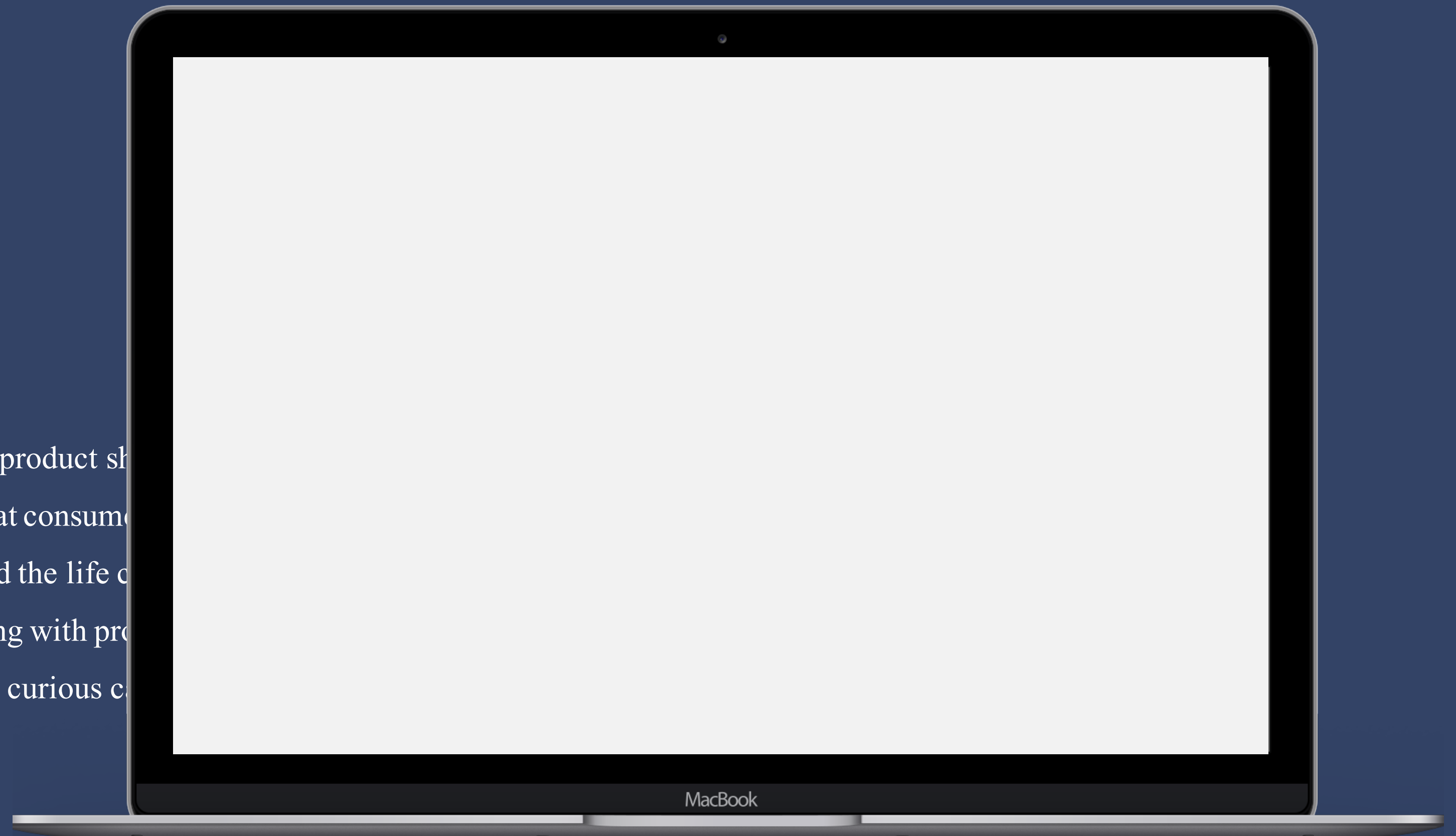
Refers to a good or service being offered by a company. Ideally, a product should meet a certain consumer demand, or it should be so compelling that consumers believe they need it. To be successful, marketers should understand the life cycle of a product, and business executives should have a plan for dealing with products at every stage of their life cycles. Some fifty years ago there was a curious case of whale-trover litigated in England.



MacBook

MacBook

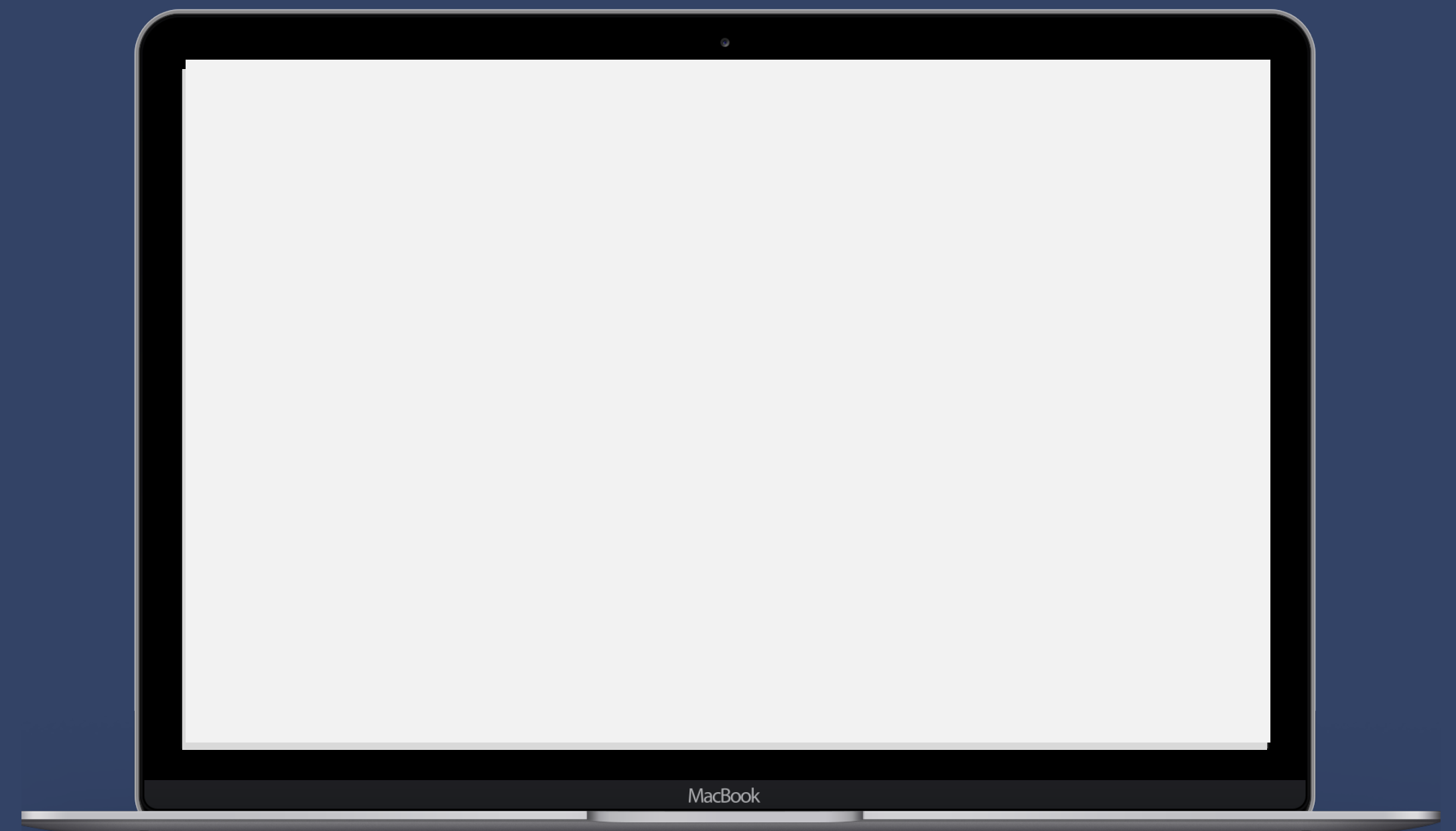
Refers to a good or service being offered by a company. Ideally, a product should meet a certain consumer demand, or it should be so compelling that consumers believe they need it. To be successful, marketers should understand the life cycle of a product, and business executives should have a plan for dealing with products at every stage of their life cycles. Some fifty years ago there was a curious case of a whale-trover litigated in England.



MacBook

MacBook

Refers to a good or service being offered by a company. Ideally, a product should meet a certain consumer demand, or it should be so compelling that consumers believe they need it. To be successful, marketers should understand the life cycle of a product, and business executives should have a plan for dealing with products at every stage of their life cycles. Some fifty years ago there was a curious case of whale-trover litigated in England.



A BIT OF OUR STORY

Write here a line of text

But what plays the mischief with this masterly code is the admirable brevity of it, which necessitates a vast volume of commentaries to expound it. First: What is a Fast-Fish? Alive or dead a fish is technically fast, when it is connected with an occupied ship or boat, by any medium at all controllable by the occupant or occupants,—a mast, an oar, a nine-inch cable.

A telegraph wire, or a strand of cobweb, it is all the same. Likewise a fish is technically fast when it bears a waif, or any other recognized symbol of possession; so long as the party waiting it plainly evince their ability at any time to take it alongside, as well as their intention so to do.

These are scientific commentaries; but the commentaries of the whale men themselves sometimes consist in hard words and harder knocks—the Coke-upon-Littleton of the fist. True, among the more upright and honorable whale men allowances are always made for peculiar cases, where it would be an outrageous moral injustice for one party to claim possession of a whale previously chased or killed by another party. But others are by no means so scrupulous.

Some fifty years ago there was a curious case of whale-trover litigated in England, wherein the plaintiffs set forth that after a hard chase of a whale in the Northern seas; and when indeed they (the plaintiffs) had succeeded in harpooning the fish; they were at last, through peril of their lives, obliged to forsake not only their lines, but their boat itself. Ultimately the defendants (the crew of another ship) came up with the whale, struck, killed, seized, and finally appropriated it before the very eyes of the

GENERAL

