EMBRACING SOCIETY 5.0

WITH HUMANITY

Editor: Diah Karmiyati

Bildung

Embracing Society 5.0 with Humanity

Embracing Society 5.0 with Humanity

Embracing Society 5.0 with Humanity

Society 5.0 is a concept presented by the Japanese as a core concept of their economic system. They believed that technology should not surpass the intelligent of men. As such, in society 5.0 the Japanese government would like to ensure that all technological things are designed to be a human-centered design. In fact, their ministry of education in 2018 has also been readily prepared the future generation through a change in their education system. For example, the minister explains that in Japan, or many parts of the world, university entrance are divided into two main concentrations, which are science and social science. The minister thought of changing the system, as society 5.0 is about creating a technology that is human centered. For instance, they gave an example on designer babies. If, people from hard science learn about philosophy, ethics, and humanities, they won't face such ethical concern when developing a product. This is what is being envision by the Japanese government for their younger generation. Collaboration between science and social science is necessary to build a better environment for our future children. Another example is the companies in Japan, such as Hitachi and Fujitsu has already been implementing this 5.0 by designing product that relied fully on technology but puts human at its center (Hitachi, 2017).

Likewise, it is currently a hot topic in Indonesia. Indonesia as a country with the 4th largest population in the world has not been implemented this concept. Our country is still on the industry 4.0. Yet, with the rising interest in AI, Blockchain, NFT, number of unicorn start-up. and all recent technological changes, our country are ready to compete with any other countries in Southeast Asia. Society 5.0 is coming, and we need to embrace it. To prepare with the society 5.0, It is not only the technological side. It is necessary for us to have a strong principle at hearts that based on our belief system. We, as an Indonesian have known to be religious that most of us believed in God. We also commonly practice our religion and tend to be kind to people because we know God would love our good deeds. I personally think that this will help us to move forward and live together with advanced technology.

Technology begets a very important leap in human's life journey. It is important to keep valued of the benefit but it's more important to look out for the human itself. As its purpose is smarter than us, to help us, it will be very ideal if we embrace the technology using our ability to be kind.

Malang, 21 Maret 2022

Diah Karmiyati

Corpus and Data-Driven Learning: Big Data for Language Teachers

Ikmi Nur Oktavianti

Introduction

Language teaching activity conducted in the EFL context relies significantly on the teaching materials and classroom interaction [1]. As a foreign language, the target language is not widely used daily. This condition itself is not adequate to provide rich and quality language input for the learners. In terms of the English language, learners in Indonesia might get exposure from many aspects due to the advance of Internet technology. However, not all learners are highly motivated; thus, teachers still hold the primary roles for their learning achievement. To help the teaching activity, teachers then use textbooks as the main resources. Surprisingly, some studies have shown that textbooks might not be the best teaching and learning resources in the EFL context [2]–[4].

Some previous studies show the differences in language presented in the textbooks and actual use of English in several aspects, such as modal verbs [3], [5]–[7]. For example, necessity/obligation modals are overused in an Indonesian EFL textbook for senior high school. Meanwhile, in COCA, prediction modals are the most frequent ones [8]. Another study also demonstrates that the use of adverb collocates of *good* and *great* in senior high school textbooks is restricted to *very*, which is inadequate [4]. Several studies also prove that textbook conversations are not natural and do not represent actual English conversations [9], [10]. Considering that the writers of the textbooks in the EFL context are mostly not native speakers of English, this phenomenon is inevitable. However, there must be an alternative solution to overcome these mismatches.

Related to the actual practice of English, it is important to be grammatically correct and naturally used. Considering that English is still a foreign language in Indonesia, this is such a challenging task for language teachers. Therefore, to ensure the quality of our language input, it is recommended to consult native speakers. However, it is nearly implausible to consult native speakers daily in language teaching activities in the EFL context. Thus, it is essential to consider using language-related technology. It is now possible to collect native speakers' use of English massively and digitally to consult this text

collection for the naturalness of 'our' English. This large-scale text collection is the manifestation of *big data* in the language context.

Big data is related to data management technology. Big data enables people to store, manage, and manipulate vast amounts of data at the right speed and time to gain the right insights [11]. There are some characteristics of big data, namely (1) a large volume of data, (2) high velocity of data, (3) a wide variety of data [11]. Since Industrial 4.0, where people start inventing the Internet and information is much easier to access, we can see the growth of big data. People in various fields, such as economics, use big data to make a comprehensive analysis of economic trends, etc. [12]. In terms of language data, it is plausible to compile various texts from various sources to a very great extent. We can have more comprehensive resources to serve language research once conducted on a smaller scale (e.g., a movie, a novel, a news article, etc.). Concerning the nature of language use as a collective practice [13], the analysis results are also more valid and representative.

More interestingly, this text collection is not only used for language research, but it is also applicable and beneficial for language teaching and learning [14], [15]. Since this collection provides many authentic examples, teachers and learners can better understand language use. This huge text collection is called *corpus*. However, the ELT practice seems to be unaffected by the emergence of corpus technology, and relatively few teachers and learners know about this valuable resource and experience the use of a corpus themselves [16]. Thus, it is crucial to introduce corpus and use it for language teaching and learning, especially in the EFL context like in Indonesia.

Discussion

Corpus (*pl. corpora*) can be defined as the large collection of texts stored digitally in machine-readable formats [17], [18]. The collection of texts is collected from naturally occurring language data. They are authentic, meaning that they are compiled from the actual use of English, such as newspapers, magazines, fiction, and academic articles. Some corpora also contain spoken language data, such as TV shows, public lectures, movies, etc. Thus, we can say that it is a representative collection of language that can be used to make statements about language use. A corpus has the characteristics of being empirical since it collects and provides actual patterns of use in natural language texts, being authentic because it compiles the language data from authentic sources, and register-based because it

differentiates spoken vs. written language data [18]. A corpus usually consists of millions or billion words that require computer software or computational mechanism to do the analysis, such as frequency analysis, concordance analysis, collocation analysis, n-gram analysis, etc. [19]. Because the size of the text collection is massive, it is therefore quite valid as a resource for language teaching and learning.

There are many English corpora available online and accessible for free. First, there are several corpora used for general purposes and needs (standard language, synchronous language), such as *Corpus of Contemporary American English* (COCA)*, British National Corpus* (BNC)*, Global Web-based English* (GloWbE), etc. There are also some specialized corpora, such as *Michigan Corpus of Upper-Level Student Papers* (MICUSP), Corpus of African American Languages (CORAAL), *Corpus of Historical American English* (COHA), to serve specific purposes, e.g., documentation of ancient languages, documentation of vernacular languages, documentation of learners' language. However, in this chapter, I focus on the use of general reference corpora for language teaching and learning due to some reasons, such as (a) the size is gigantic (thousand million or hundred million or billion words), (b) it is updated, and (c) it prioritizes standard language which is in line with the 'type' of the language taught at schools. I highly recommend using *COCA* and *BNC* as the consultation corpus for teaching and learning activities.

Corpus for Language Teachers

The use of corpus in English language teaching is not new. Corpus technology has been used as a basis for teaching and learning languages for more than 25 years. One of the earliest corpora was in the 1980s when John Sinclair (professor in the Department of English) collaborated with Collins, publishing on the *COBUILD* project in pedagogically oriented lexical computing [16]. A much earlier attempt to improve English language teaching by collecting authentic sources was the *General Service List* (GSL) compiled by Michael West [16]. This list presents the most frequently used words in the English language, which was determined based on empirical bases. It is not surprising that all major learners' dictionaries of English are now based on constantly updated multi-million-word language databases. Not only dictionaries, but some coursebooks are also now written based on the results of corpus investigation or corpus consultation [20].

There are two types of corpus applications for language teaching and learning. The first is the indirect use of corpus in the

classroom which means corpus is used as a teaching and learning media and resource. Despite being used directly, corpus can also be used indirectly, such as shown in figure 1 [16].
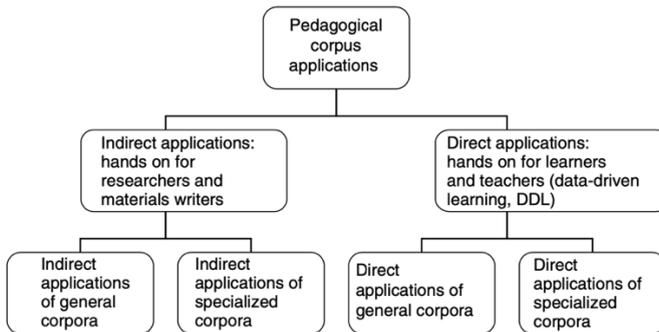

Figure 1. Corpus use for language teaching and learning

This chapter focuses on using a corpus directly in the classroom, which is known as *Data-Driven Learning.*

Data-Driven Learning

An attempt to minimize the discrepancies or mismatches of languages we teach and the actual use of English can be made by using a corpus as a teaching media. Corpus can be used directly in the classroom for practicing an inductive approach. Data-Driven Learning (DDL) is the activity of using a vast language database to explore the structure, pattern, meaning, etc. [14], [21]. Tim Johns first introduced this term when working with parts of the *COBUILD* corpus [21]. There are some applications of corpus use in the DDL practice. Teachers can use a corpus, for instance, COCA, to check the most commonly used grammatical patterns and involve the students in this activity. Here is an example of investigating the prepositional collocate for *interested.*

**Figure 2. Results of corpus investigation (*interested +
prep*)**

Figure 2 demonstrates the natural combination for *interested* + prep is *interested in* having the highest frequency of use in the COCA and leaving the rest as minor variants. By doing this activity, students can experience the investigation themselves and develop their inductive thinking. For the full contexts, COCA can also display the use of *interested in* as in figure 3.

Figure 3. Examples of *interested in* from COCA

Furthermore, students can also observe the examples provided by the corpus and identify how they are used. From these examples, students can also generate patterns. This activity strongly supports discovery learning and makes students more active in the learning process.

Another corpus application in the classroom is for meaning identification. The word *common,* for example, conveys several meanings. This word can mean differently depending on the context,

and a corpus can be the perfect tool to identify which meaning belongs to which context of use. Below are some possible meanings of *common* from the Oxford Dictionary.

com·mon | ˈkämən |

adjective (**commoner, commonest**)

1 occurring, found, or done often; prevalent: *salt and pepper are the two most common seasonings* | *it's* **common for** *children to have middle ear infections.*
- (of an animal or plant) found or living in relatively large numbers; not rare: *you might spot less common birds such as the great spotted woodpecker* | *the swordfish is not common in European waters.*
- denoting the most widespread or typical species of an animal or plant: *the common blue spruce.*
- ordinary; of ordinary qualities; without special rank or position: *the dwellings of common people* | *a common soldier.*
- (of a quality) of a sort or level to be generally expected: *common decency.*
- of the most familiar type: *the common or vernacular name.*

2 shared by, coming from, or done by more than one: *the two republics' common border* | *problems* **common to** *both communities.*
- belonging to, open to, or affecting the whole of a community or the public: *common land.*
- *Mathematics* belonging to two or more quantities.

3 *British* showing a lack of taste and refinement; vulgar: *she's so common.*

4 *Grammar* (in Latin and certain other languages) of or denoting a gender of nouns that are conventionally regarded as masculine or feminine, contrasting with neuter.
- (in English) denoting a noun that refers to individuals of either sex (e.g. *teacher*).

5 *Prosody* (of a syllable) able to be either short or long.

6 *Law* (of a crime) of relatively minor importance: *common assault.*

Figure 4. Meanings of *common* (adj)

Teachers can direct the students to use the KWIC feature in COCA to observe the use of 'common' from authentic sources, as shown in figure 5. Teachers can assign the students to identify the meaning of *common* in the concordance lines

Figure 5. Some concordance lines of *common*

DDL with a corpus can also be done to deal with synonyms. The words *pretty* and *beautiful* are synonymous. However, synonyms do not mean they have the same meaning, and they are not always interchangeable. Therefore, it is essential to equip the students with sufficient knowledge. With the assistance of a corpus, teachers can provide a more useful and reliable resource to tell the difference. In COCA, for example, there is a feature to compare synonyms. Here is the result for *pretty* and *beautiful.*

WORD 1 (W1): **PRETTY** (1.84)						WORD 2 (W2): **BEAUTIFUL** (0.54)					
	WORD	W1	W2	W1/W2	SCORE		WORD	W2	W1	W2/W1	SCORE
1	MUCH	31809	0	63,618.0	34,542.0	1	FRIENDSHIP	103	0	206.0	379.4
2	SURE	8749	0	17,498.0	9,500.7	2	KIDS	101	0	202.0	372.0
3	SOON	3259	0	6,518.0	3,539.0	3	EXPERIENCE	82	0	164.0	302.0
4	GOOD	22863	6	3,810.5	2,068.9	4	HOMES	81	0	162.0	298.4
5	AMAZING	1368	0	2,736.0	1,485.5	5	MOMENTS	55	0	110.0	202.6
6	OBVIOUS	1347	0	2,694.0	1,462.7	6	SPIRIT	54	0	108.0	198.9
7	TOUGH	1153	0	2,306.0	1,252.1	7	TRIBUTE	54	0	108.0	198.9
8	DARN	1030	0	2,060.0	1,118.5	8	SOUL	106	1	106.0	195.2
9	FUNNY	979	0	1,958.0	1,063.1	9	FORM	48	0	96.0	176.8
10	EASY	1577	1	1,577.0	856.2	10	SON	95	1	95.0	175.0

Figure 6. Collocates of *pretty* and *beautiful* in COCA

Figure 6 presents the collocates of *pretty* and *beautiful* in COCA, and it is evident that they are used differently. The word *pretty* is more frequently used as an adverb to modify adjectives (e.g., *good, sure, amazing, obvious, funny*, etc.). In contrast, the word *beautiful* is an adjective to modify nouns (*friendship, kids, experience, homes, moments*, etc.). In other words, *pretty* is used more frequently as an adverb which makes it not interchangeable with *beautiful* to all extents. Teachers and learners can maximize the learning experience and better understand the target language by having empirical and massive data.

The teachers might also simplify DDL experience. The raw results of corpus search can be extracted and selected due to learners' proficiency and needs. Below is an example of the materials on necessity/obligation modals, *must* and *should,* taken from a corpus and adjusted into a more straightforward presentation in figure 7.

1	said, anybody who is interested in becoming a (unclear) in nineteen ninety three,	**must**	register an interest on sixteenth this year, which is really
2	He should know about them shouldn't he? (unclear). Well we	**must**	(pause) that's why, certainly have to, before (unclear)
3	A B C whenever they are put in these cardboard boxes they have to be sorted so somebody	**must**	sort them.
4	A B C're not asking (unclear) sorting, oh I see. Erm (pause) so there	**must**	be somebody responsible and I think if, if we write swimming pool and people
5	A B C collecting cost. Because the (pause) money they can get for this (pause) mixed paper	**must**	be a pittance. Mm. Almost hardly worth their while but (unclear)
6	A B C Charity shops. (Ah (pause) er the other item that we	**must**	mention is any extra plastic bags Yeah, I'll ask about plastic bags
7	what see says any rate on Monday. But I don't think we	**should**	give her too long, cos she's got herself into (unclear)2. Mm.
8	A B C at Kathy (-----) says, she's afraid of being left out, if there	**should**	be a, another Tory Government, so a lot of people are in the
9	A B C it and to say how, how they tackled it, so I think it	**should**	be an interesting meeting, hopefully we'll get a few, you know interesting people
10	A B C women's committee and they assure me that they, the hospital says these tests	**should**	come through in a week, at the most three week'
11	A B C if we established a, a demand, and I don't see why we	**shouldn't**	have a bash at. Well then I have the inspector (unclear)
12	A B C S A branch, you know, I don't see why, why we	**shouldn't**	give it a try. There is one at the (unclear)

Figure 7. Concordance lines of *must* and *should*

The examples provided in figure 7 are authentic and natural, which can support the students in understanding the actual use of *must* and *should.* The application of DDL using a corpus or some corpora might be beneficial for vocabulary teaching [22] and grammar teaching [23].

The Benefits and the Drawbacks of Corpus

Corpus in language teaching and learning offers some benefits since they are based on actual language use, meaning that teachers do not invent the examples [20]. The sources of the examples are authentic, which guarantee the expressions or word combinations or patterns are naturally used in the English language. Moreover, using a corpus can distinguish between spoken and written languages because both have different characteristics and distinctive linguistic features. Thus, by accessing and using a corpus or corpora, the students don't have to live in the target language environment to experience the actual use of English [20]. Considering the size, a corpus can also offer many examples for teachers and learners, later using them as parts of teaching materials or test items. Thus, the corpus size is one of the strengths of a corpus that we hardly find in any other teaching and learning resources.

Apart from the strengths of corpus, which provide some benefits for language teachers, there are some drawbacks of using corpus. Apart from the large-scale data provided by a corpus, it informs us whether the pattern has occurred and whether it is

frequent [24], [25], but it cannot inform us about what is possible and is not related to grammaticality or ungrammaticality. Regarding the application of corpus, the findings of corpus search or corpus activity should be adjusted to the needs of the learners and the course [20]. The results of corpus investigation or corpus consultation sometimes cannot be used directly due to the learners' proficiency level. It leads to the demand for another skill, which also needs extra effort from the teachers who use the corpus.

Furthermore, it is also important to consider the technical aspects of using the corpus. Since corpus is quite technical, to use it, teachers should be familiar with it. Some terms are new for language teachers, especially if they do not have strong knowledge of linguistics. These terms, such as *concordance, collocation, KWIC, PoS Tag, collocation span, n-gram*, and others, are pretty much confusing for new users of a corpus. Besides, we should have a good Internet connection to smoothly access the corpus, which can be a constraint for some teachers and students in remote areas.

Conclusion

As a tool to help teach and learn English, corpus offers many benefits. As a teaching and learning media and resource, corpus can be accessed and used directly in the classroom to assist both teachers and learners in getting quality language input. Teachers should highly consider this activity because the intuition of a non-native speaker of English is not a reliable source. In the Society 5.0 era, the Internet is tremendously faster, and big data continues to be very useful and prominent in various aspects of life, including language teaching. If there is a huge source of English language data we can use and consult, how can we just ignore it? We can make use of corpus and take some benefits. With regard to notable benefits of corpus, it can be the best companion to the existing textbooks used by teachers. However, corpus applications should be equipped with sufficient training, which means corpus-related courses should be taught in the teaching department. Good internet connection should also be provided for more comfortable teaching and learning sessions.

References
[1]   I. McGrath, *Teaching Materials and the Roles of EFL/ESL Teachers: Practice and Theory*. London: Bloomsbury, 2013.
[2]   U. Römer, "A Corpus-Driven Approach to Modal Auxiliaries and their Didactics," in *How to Use Corpora in Language Teaching*, J.

M. Sinclair, Ed. Amsterdam: John Benjamin Publishing Company, 2004.

[3] J. Mukundan and L. Khojasteh, "Modal Auxiliary Verbs in Prescribed Malaysian English Textbooks," *ELT*, vol. 4, no. 1, p. 79, Feb. 2011, doi: 10.5539/elt.v4n1p79.

[4] I. N. Oktavianti and J. Sarage, "Collocates of 'great' and 'good' in the Corpus of Contemporary American English and Indonesian EFL Textbooks," *SiELE*, vol. 8, no. 2, pp. 457–478, May 2021, doi: 10.24815/siele.v8i2.18594.

[5] J. Holmes, "Doubt and Certainty in ESL Textbooks," *Applied Linguistics*, vol. 9, no. 1, pp. 21–44, Mar. 1988, doi: 10.1093/applin/9.1.21.

[6] D. Shin and Y. V. Chon, "A Corpus-based Analysis of Curriculum-based Elementary and Secondary English Textbooks," *Multimedia-Assisted Language Learning*, vol. 14, no. 1, pp. 145–179, Apr. 2011, doi: 10.15702/MALL.2011.14.1.145.

[7] A. Molavi, M. Koosha, and H. Hosseini, "A Comparative Corpus-Based Analysis of Lexical Collocations Used in EFL textbooks," *LACLIL*, vol. 7, no. 1, pp. 66–81, Apr. 2014, doi: 10.5294/laclil.2014.7.1.4.

[8] I. N. Oktavianti and A. Fajria, "Modal Verbs in a Curriculum-Based EFL Textbook of Senior High School in Indonesia: A Corpus-Based Study," *TESOL International Journal*, vol. 16, no. 5, pp. 171–204, 2021.

[9] I. N. Oktavianti, I. Prayogi, M. A. Amal B., and R. S. Pertiwi, "An Analysis of Conversations in Curriculum-based EFL Textbooks for Senior High School in Indonesia and the Comparison with Corpus-based English Textbooks," *Universal Journal of Educational Research*, vol. 8, no. 9, pp. 4151–4162, Sep. 2020, doi: 10.13189/ujer.2020.080941.

[10] A. Gilmore, "A Comparison of Textbook and Authentic Interactions," *ELT Journal*, vol. 58, no. 4, pp. 363–374, Oct. 2004, doi: 10.1093/elt/58.4.363.

[11] J. Hurwitz, A. Nugent, F. Halper, and M. Kaufman, *Big Data for Dummies*. New Jersey: John Wiley & Sons, Inc., 2013.

[12] M. Kitsuregawa, "Transformational Role of Big Data in Society 5.0," in *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, Dec. 2018, pp. 3–3. doi: 10.1109/BigData.2018.8621989.

[13] A. Burkette and W. A. Kretzschmar Jr., *Exploring Linguistic Science: Language Use, Complexity, and Interaction*, 1st ed.

Cambridge University Press, 2018. doi: 10.1017/9781108344326.

[14] I. Timmis, *Corpus Linguistics for ELT*, 0 ed. London: Routledge, 2015. doi: 10.4324/9781315715537.

[15] A. O'Keeffe, M. Mccarthy, and R. Carter, *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press, 2007.

[16] U. Römer, "Using General and Specialized Corpora in English Language Teaching: Past, Present and Future," in *Corpus-Based Approaches to English Language Teaching*, M. C. Campoy, M. L. Gea-Valor, and B. Belles-Fortuño, Eds. London: Continuum, 2010.

[17] T. McEnery and A. Hardie, *Corpus Linguistics*. Cambridge: Cambridge University Press, 2012.

[18] W. Crawford and E. Csomay, *Doing Corpus Linguistics*, 0 ed. Routledge, 2015. doi: 10.4324/9781315775647.

[19] P. Szudarski, *Corpus Linguistics for Vocabulary: A Guide for Research*, 1st ed. London: Routledge, 2017. doi: 10.4324/9781315107769.

[20] M. McCarthy, *Touchstone: From Corpus to Course Book*. New York: Cambridge University Press, 2004.

[21] A. Boulton, "What Data for Data-Driven Learning?" *Eurocall*, vol. 20, no. 1, p. 36, Mar. 2012, doi: 10.4995/eurocall.2012.16038.

[22] P. Liu, "Data-Driven ESP Vocabulary Learning," in *International Conference on Modern Education and Social Science*, 2016, pp. 219–225.

[23] S. I. Nugraha, F. Miftakh, and K. Wachyudi, "Teaching Grammar through Data-Driven Learning (DDL) Approach," presented at the Ninth International Conference on Applied Linguistics (CONAPLIN 9), Bandung, Indonesia, 2017. doi: 10.2991/conaplin-16.2017.68.

[24] S. Hunston, *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press, 2002.

[25] L. Flowerdew, *Corpora and Language Education*. London: Palgrave Macmillan UK, 2012. doi: 10.1057/9780230355569.

EMBRAC!NG SOCIETY 5.0 WITH HUMANITY

Society 5.0 is a concept presented by the Japanese as a core concept of their economic system. They believed that technology should not surpass the intelligent of men. As such, in society 5.0 the Japanese government would like to ensure that all technological things are designed to be a human-centered design. In fact, their ministry of education in 2018 has also been readily prepared the future generation through a change in their education system. For example, the minister explains that in Japan, or many parts of the world, university entrance are divided into two main concentrations, which are science and social science. The minister thought of changing the system, as society 5.0 is about creating a technology that is human centered. For instance, they gave an example on designer babies. If, people from hard science learn about philosophy, ethics, and humanities, they won't face such ethical concern when developing a product. This is what is being envision by the Japanese government for their younger generation. Collaboration between science and social science is necessary to build a better environment for our future children. Another example is the companies in Japan, such as Hitachi and Fujitsu has already been implementing this 5.0 by designing product that relied fully on technology but puts human at its center (Hitachi, 2017).

Likewise, it is currently a hot topic in Indonesia. Indonesia as a country with the 4th largest population in the world has not been implemented this concept. Our country is still on the industry 4.0. Yet, with the rising interest in AI, Blockchain, NFT, number of unicorn start-up, and all recent technological changes, our country are ready to compete with any other countries in Southeast Asia. Society 5.0 is coming, and we need to embrace it. To prepare with the society 5.0, It is not only the technological side. It is necessary for us to have a strong principle at hearts that based on our belief system. We, as an Indonesian have known to be religious that most of us believed in God. We also commonly practice our religion and tend to be kind to people because we know God would love our good deeds. I personally think that this will help us to move forward and live together with advanced technology.

Technology begets a very important leap in human's life journey. It is important to keep valued of the benefit but it's more important to look out for the human itself. As its purpose is smarter than us, to help us, it will be very ideal if we embrace the technology using our ability to be kind.