CANARA ENGINEERING COLLEGE
DEPARTMENT OF INFORMATION SCIENCE
Benjanapadavu, Bantwal Taluk, Mangalore – 574219

Department of Information Science and Engineering

VISION

The Department of Information Science and Engineering strives to be a centre of learning in the field of Information Technology to produce globally competent engineers catering to the needs of the industry and society.

MISSION

- Impart technical skills in the field of Information Science & Engineering.
- Train and transform students to become technological thinkers and facilitate a quality venture which meets the industrial and societal needs.
- Encourage students to become well-rounded in their professional competencies.

PROGRAM EDUCATIONAL OBJECTIVES

1. Graduates will succeed in the field of Information Science and Engineering, professional career and higher studies.
2. Graduates will analyze the requirements of the software industries and provide novel engineering designs and efficient solutions with legal and ethical responsibility.
3. Graduates will adapt to emerging technologies, work in multidisciplinary teams with effective communication skills and leadership qualities.

PROGRAM OUTCOMES

Engineering graduates in **Information Science and Engineering**will be able to:

1. **Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. **Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and Sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication:** CCommunicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
12. **Life-long learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAM SPECIFIC OUTCOMES

1. An ability to understand, analyze and impart the basic knowledge of Information Science and Engineering.
2. An ability to apply the programming, designing, and problem solving techniques in building/simulating the applications, solving the problems and guiding the innovative career paths to become an IT Engineer.
3.

COURSE LEARNING OBIECTIVES OF RPA:

This course(18CS745) will enable students to:

1. To understand basic concepts of RPA
2. To Describe IIPA, where it can be applied and how it implemented
3. To Describe the different types of variables, Control Flow and data manipulation techniques
4. To Understand Image,Text lnd nata Tables Automation
5. To Describe various types ofExceptions and strategies to handle

COURSE OUTCOMES:

The students should be able to :

6. To Understand the basic concepts of RPA
7. To Describe various components and platforms of RPA
8. To Describe the different types of variables, control flow and data manipulation techniques
9. To Understand various control techniques and OCR in RPA
10. To Describe various types and strategies to handle exception

TEXT BOOKS:

1. Tom Taulli, The Robotic Process Automation Handbook:A Guide to Implementing RPA Systems,2020,lSBN-13 (electronic):978-7-4842-5729-6, Publisher : A press
2. Alok Mani Tripathi, Learning Robotic Process Automation, Publisher: Packt Publishing Release Date: March 2018 ISBN: 9787788470940

MODULE 1

RPA FOUNDATIONS

WHAT IS RPA?

RPA (Robotic Process Automation) is the technology that enables software 'robots' to carry out repetitive, rule-based digital tasks. Humans typically perform these tasks through the user interface, using the mouse and keyboard. RPA robots are capable of mimicking the human actions, and they are typically more accurate, faster, and more consistent at it.

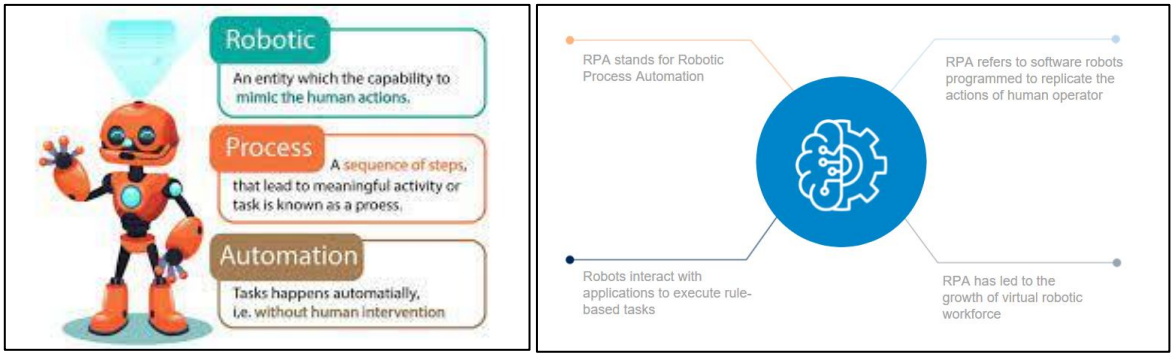


Fig.1.1: Definition of Robotic Process Automation

Combined with Artificial Intelligence (AI), RPA can target more sophisticated work. Automation is a term that describes more accurately these possibilities that exceed the sphere of basic RPA. Sometimes we use these two terms interchangeably, given that RPA is still at the core of automation. The word “robotic” does not refer to a physical robot – instead, it is about a software-based robot (or bot) that can automate human actions in the workplace. A bot can be delivered via the cloud or through downloadable software. However, the use of robotic does look like a savvy marketing move. Even the word “process” is not particularly descriptive either. A better alternative would be “tasks,” which are individual action items that are a part of a process. RPA involves bots that perform a set of specified actions or tasks, such as the following:

- The cut-and-paste of information from one app to another
- The opening of a web site and login
- The opening of an e-mail and attachments
- The read/write of a database
- The extraction of content from forms or documents
- The use of calculations and workflows

FLAVORS OF RPA

There are different types of RPA approaches. Part of this is due to the fact that the technology is continuing to evolve. Vendors are also looking at ways to redefine RPA so as to help them stand out in the marketplace. On a high level, you can divide the flavors into the following:

1. **Attended RPA** (which may be referred to as robotic desktop automation or RDA): This was the first form of RPA that emerged, back in 2003 or so. Attended RPA means that the software provides collaboration with a person for certain tasks. A prime example would be in the call center, where a rep can have the RPA system handle looking up information while he or she talks to a customer.
2. **Unattended RPA**: This technology was the second generation of RPA. With unattended RPA, you can automate a process without the need for human involvement – that is, the bot is triggered when certain events happen, such as when a customer e-mails an invoice. Consider that unattended RPA is generally for back-office functions.
3. **Intelligent process automation or IPA** (this may also be referred to as cognitive RPA): This is the latest generation of RPA technology, which leverages AI to allow the system to learn over time (an example would be the interpretation of documents, such as invoices). Because of this, there may be even less human intervention since the RPA software will use its own insights and judgements to make decisions.

It's important to understand these variations because some RPA systems may specialize in a particular approach. Besides, when looking at your own needs for automation, it's a good idea to see what types may work the best

HISTORY OF RPA

No doubt, the concept of automation is far from new. Yet it would not be centuries later until notable real-world examples of automation would emerge. After all, it's only been during the past 70 or so years that computers have been a major catalyst for this trend. Along the way, there have been different periods of automation, based on the types of technologies available. They would also provide a foundation for RPA platforms.

1. **Mainframe Era:** These were huge machines developed by companies like IBM. They were expensive and mostly available to large companies. Yet they were incredibly useful in helping manage core functions for companies, such as payroll and customer accounts.
2. **PC Revolution**: Intel's development of the microprocessor and Microsoft's development of its operating system revolutionized the technology industry. As a result, just about any business could automate processes, say by using word processors and spreadsheets.

But the automation technologies – while powerful – still had their drawbacks. They could easily result in complex IT environments, which required expensive and time consuming integrations and custom coding. Because of this, an employee may have to use multiple applications in their daily activities that could involve wasteful tasks like moving data from one to the other. The irony was that the technology could make employees less productive! From this emerged the key elements for RPA, which came about in the early 2000s. A big part of this was screen scraping, which is the automation of moving data among applications, which turned out to provide a nice boost to efficiency and effectiveness. But the nascent RPA market got scant attention. It was mostly perceived as low-tech and a commodity. Instead, investors and entrepreneurs in Silicon Valley focused their attention on the rapidly growing cloud market that was disrupting traditional IT systems.

But around 2012 or so, the RPA market hit an inflection point. There was a convergence of trends that made this happen, such as the following:

- In the aftermath of the financial crisis, companies were looking for ways to lower their costs. Simply put, traditional technologies like ERP were reaching maturation. So companies needed to look for new drivers.
- Companies also realized they had to find ways to not be disrupted from technology companies. RPA was considered an easier and more cost-effective way to go digital.
- Some industries like banking were becoming more subject to regulation. In other words, there was a compelling need to find ways to lessen the paperwork and improve audit, security, and control.
- RPA technology was starting to get more sophisticated and easier to use, allowing for higher ROI (return on investment).
- Large companies were starting to use RPA for mission-critical applications.
- Demographics were also key. As the millennials started to enter the workforce, they wanted more engaging work. They wanted careers, not jobs.

"The evolution of the RPA market is like any major technology trend," said Mihir Shukla, who is the CEO and cofounder of Automation Anywhere. "There was a gradual progress, which involved periodic breakthroughs. A prime example is the iPhone. Before this, there was a long period of incremental innovation." Fast forward to today, RPA is the fastest growing part of the software industry. According to Gartner, the spending on this technology jumped by 63% to $850 million in 2018 and is forecasted to reach $1.3 billion by 2019. Or consider the findings from Transparency Market Research. The firm projects that the global market for RPA will

soar to $5 billion by 2020. Here are some other metrics to note: By 2020, RPA along with AI will reduce the business shared-service centers by 65% (Gartner). There will also be adoption by 40% of large enterprises, compared to 10% in 2019. Based on current projections, there will likely be saturation in the RPA market by 2023 (Deloitte). The financial impact from RPA could hit $6.7 trillion by 2025 (McKinsey & Company). In terms of the global market share for RPA software, North America represents 51% and Western Europe is at 23%. But Asia is starting to get traction, especially Japan. By 2023, the forecast is that there will be $12 billion in spending on RPA services (Forrester).

THE BENEFITS OF RPA

Fig.1.2: Benefitis of Robotic Process Automation

1. **Higher quality services, greater accuracy**: With reduced human error and greater compliance, the quality of work is much better.
2. **Improved analytics**: Since these software Robots can log each action taken with the appropriate tag and metadata, it is very easy to get business insights and other analytical data. Using analytics on the collected data such as transaction received time, transaction complete time, and predictions can be made for the incoming volume and ability to complete the tasks on time.
3. **Reduced costs:** Nowadays, it is commonplace to hear that one Robot is equivalent to three human full-time executives (FTE). This is based on the simple fact that one FTE works for eight hours a day, while a Robot can work for 24 hours without a break. Increased availability and productivity means the cost of operations is reduced tremendously. The speed of the work being performed coupled with multitasking results in further reductions in cost.
4. **Increased speed**: Robots are very fast and sometimes the speed of execution has to be reduced to match the speed and latency of the application on which these Robots work. Increased speed can result in better response times and an increase in the volume of the tasks being performed.

5. **Greater compliance**: As mentioned earlier, a full audit trail is one of the highlights of RPA and can result in greater compliance. These Robots do not deviate from the defined set of steps to be taken while doing a task and hence it will certainly result in better compliance.
6. **Agility**: Reducing and increasing the number of Robot resources requires managing the volume of the business process. This is just a click away. More Robots can be deployed to perform the same task easily. Redeployment of resources does not require any kind of coding or reconfiguration.
7. **Comprehensive insights:** In addition to the audit trail and time stamping, Robots can tag transactions to use them later, in reports for business insight. By using these business insights, better decisions can be made for the improvement of the business. This recorded data can also be used for forecasting.
8. **Versatility**: RPA is applicable across industries performing a wide range of tasksdfrom small to large businesses, simple to complex processes.
9. **Simplicity**: RPA does not need prior programming knowledge. Most platforms provide designs in the form of flowcharts. This simplicity enables easy automation of business processes, leaving the IT professionals relatively free to carry out higher value work.
10. **Scalability**: RPA is highly scalable, up as well as down. Whether one requires an increase or reduction in the virtual workforce, Robots can be quickly deployed at zero or minimum costs while maintaining consistency in the quality of work
11. **Time savings**: Not only does the virtual workforce complete large volumes of work in a shorter span of time with precision, but they help save time in another way too. If there is any changedsay, a technology upgrade it is much easier and faster for the virtual workforce to adapt to the changes. This can be done by bringing about modifications in the programming or introducing new processes. For humans, it is difficult for them to learn and get trained in something newdbreaking from the old habit of performing repetitive tasks.
12. **Increased employee satisfaction**: With repetitive, dreary tasks now being taken over by the virtual workforce, employees are not just relieved of their workload, but can also engage in better quality work that requires the use of human capabilities and strengths such as emotional intelligence, reasoning, or tending to customers. Thus, RPA doesn't take away work, it just frees humans from tedious, mind-numbing work, giving us an opportunity to engage in much more satisfying jobs.

THE DOWNSIDES OF RPA

Though there are advantages to RPA, there are some significant drawbacks. Here are some issues you should look out for:

1. **Long-term sustainability.** RPA can become a serious decoy from the necessary long-term work needed to digitize and make processes and administrative work more efficient. There is a risk that you may focus on quick fixes rather than doing things the correct way from the start.
2. **Implementation**. RPA might give you good value in stand-alone tasks, but it takes time and money to set up. It's a bad idea to cut corners on RPA, too, since a miscalibration will lead to errors. An inefficient process won't be transformed simply because you add some "smartness" at the top. You should know that almost half of all RPA systems fail when first rolled out, so the implementation might take more resources than you initially thought.
3. **Error magnification**. RPA robots can't detect some obvious errors that a human would be able to immediately point out. If your data has problems with it, RPA robots will not call it out, but pass it on, magnifying an error that might have otherwise been caught.
4. **Overall risk**. Some problems aren't a good fit for RPA, especially when the stakes are high. For example, if you need to handle your purchase invoices, it's likely a better idea to use software that is able to understand and manage the data correctly from the start.
5. **Maintenance**. Most RPA solutions have to be custom-made to fit your business. It likely won't be worth it to invest in such a system if the way your business runs could change drastically in the future. Even minor changes in your setup can create significant disruption for your RPA robots.

RPA COMPARED TO BPO, BPM AND BPA

1. BPM (Business Process Management)

With the intense competition from Japan during the 1970s and 1980s, US companies were desperately seeking new and innovative approaches to improve their efficiency and competitiveness. Part of this meant adopting different management approaches, such as Six Sigma (this includes a combination of project management and statistical techniques), lean production (which is based on the manufacturing principles of Toyota), and total quality management or TQM (a blend of Six Sigma and lean production). There was also a greater focus on computer technologies. For example, FileNet introduced a digital workflow management system to help better handle documents (the company would eventually be

purchased by IBM). Then there would come onto the scene ERP vendors, such as PeopleSoft. All of this would converge into a major wave called BPM. For the most part, the focus was on having a comprehensive improvement on business processes. This would encompass both optimizing systems for employees but also IT assets.

With BPM, it requires much more time and effort with the implementation because it is about changing extensive processes, not tasks. There also needs to be detailed documentation and training. Because of this rigorous approach, BPM is often attractive to industries that are heavily regulated, such as financial services and healthcare. However, the risk is that there may be too much structure, which can stifle innovation and agility. On the other hand, RPA can be complementary to BPM. That is, you can first undergo a BPM implementation to greatly improve core processes. Then you can look to RPA to fill in the gaps.

2. BPO (Business Process Outsourcing)

This is when a company outsources a business service function like payroll, customer support, procurement, and HR. Generally speaking, a BPO will have three types of strategies:

a) **Offshore**: This is where the employees are in another country, usually far away.
b) **Nearshore**: This is when the BPO is in a neighboring country. True, there are usually higher costs but there is the benefit of being able to conveniently visit the vendor. This can greatly help with the collaboration.
c) **Onshore**: The vendor is in the same country. For example, there can be wide differences in wages in the United States.

But as with anything, there are drawbacks with a BPO. Yet here are some others to consider:

a) **Security**: If a BPO company is developing an app with your company's data, are there enough precautions in place so there is not a breach? Even if so, it can still be difficult to enforce and manage.
b) **Costs**: Over the years, countries like China and India have seen rising labor costs. This has resulted in companies moving to other locations, which can be disruptive and expensive.
c) **Politics**: This can be a wildcard. Instability can easily mean having to abandon a BPO operator in a particular country.

3. BPA (Business Process Automation)

This is the use of technology to automate a complete process. One common use case is onboarding. For example, bringing on a new employee involves many steps, which are repeatable and entail lots of paperwork. For a large organization, the process can be time-consuming and expensive. But BPA can streamline everything, allowing for the onboarding at scale.

CONSUMER WILLINGNESS FOR AUTOMATION

The automation of consumer-facing activities, such as with chatbots on a smartphone or web site, are becoming more ubiquitous. But this begs the question: Does this technology really provide a good experience? Might it be doing more harm than good? These are very good questions to ask as many automation technologies are far from perfect, especially those that deal with complex environments. Consider a report from Helpshift, an AI-based digital customer service platform automating 80% of customer support issues for huge D2C (direct-to-consumer) brands including companies like Flipboard, Microsoft, Tradesy, and 60 others. Its report is based on the analysis of 75 million customer service tickets and 71 million bot-sent messages. Here are some of the findings: A total of 55% of the respondents – and 65% of millennials – prefer chatbots with customer service so long as it is more efficient and reduces phone time to resolve an issue and explain a problem. A total of 49% say they appreciate the 24/7 availability of chatbots. Granted, there is much progress to be made. Chatbot technology is still in the early phases and can be glitchy, if not downright annoying in certain circumstances. But in the years to come, this form of automation will likely become more important – and also a part of the RPA roadmap. According to the CEO of Helpshift, Linda Crawford: "Seeing as the vast majority of Americans dread contacting customer support, there's a huge opportunity here for chatbots to fill the void and improve the customer support experience for consumers—and agents".

THE WORKFORCE OF THE FUTURE

The modern concept of work goes back to Henry Ford. To scale his operations to create the automotive industry, he realized he had to be competitive with wages so as to attract the best employees. He made the remarkable move of doubling them $5 a day, which actually had the consequence of greatly expanding the middle class! But he also introduced other policies, such as the 40-hour workweek, with weekends off. The interesting thing is that the fundamentals of work have not changed much since then. True, there has been the trend of the gig economy, in which people get paid for offering services through Uber and Lyft. Yet when it comes to office work, the structure has remained quite durable.

But the fact remains that many of the activities remain fairly tedious and uninteresting. According to research from the McKinsey Global Institute, white collar workers still spend 60% of their time on manual tasks, such as with answering e-mails, using spreadsheets, writing notes, and making calls. Interestingly enough, a report from David White estimates that $575 billion is wasted in the United States because of inefficient processes. In light of all this, RPA

is likely to have a significant impact on the workplace because more and more of the repetitive processes will be automated away. One potential consequence is that there may be growing job losses. A survey from Forrester predicts that – as of 2025 – software automation will mean the loss of 9% of the world's jobs or 230 million. Then again, the new technologies and approaches will open up many new opportunities. Yet this may not be enough to make up for the shortfall. The Forrester study, for example, forecasts that there will be replacement of 16% of US jobs and the creation of 9% of new ones. Or look at the research from McKinsey & Company. Its analysis shows that technologies like RPA could automate a whopping 45% of the activities of a company's workforce. Now when a company engages in an automation project, the CEO will usually not talk about job loss. It's something that will frighten the workforce and generate awful headlines. Instead the messaging will be vague, focusing on the overall benefits of the transformation. This may make it sound like not much is happening. But it does seem like a good bet that the reverberations will grow and grow, as RPA systems get increasingly robust. As we've seen in prior periods where technology resulted in job loss – such as in the Industrial Revolution – there are serious changes in politics and regulations. The upshot is that technology often becomes a target of scorn.

Companies really do try to avoid layoffs, since they are expensive and take a toll on the organization. But in the years ahead, managers will probably need to find ways to navigate the changes from automation, such as finding new roles or reskilling the workforce. All in all, the rise of automation has the potential for leading for a much better society. Again, workers can focus on more interesting and engaging activities – not repetitive and mundane tasks. There will also be ongoing renewing of knowledge and understanding. But there must be proactive efforts, say from companies and governments, to provide for a smoother transition. A survey from PEGA shows that nearly 70% of the respondents believe that the concept of "workforce" will include employees and machines and 88% believe this is fine, so long as they are not managed by machines

RPA SKILLS: ON-PREMISE VS. THE CLOUD

The traditional IT system approach is the use of on-premise technology. This means that a company purchases and sets up its own hardware and software in its own data center. Some of the benefits include:

1. A company has complete control over everything. This is particularly important for regulated industries that require high levels of security and privacy.

2. With on-premise software, you may have a better ability to customize the solution to your company's unique needs and IT policies.

However, the on-premise technology model has serious issues as well. One of the biggest is the cost, which often involves large up-front capital expenses. Then there is the ongoing need for maintenance, upgrades, and monitoring. All in all, it means that the IT department may be spending valuable time on noncore activities. And finally, the use of point applications like Excel can lead to a fragmented environment, in which it becomes difficult to centralize data because there are so many files spread across the organization. Because of all this, companies have been looking at another approach – that is, *cloud computing*.

The interesting thing is that this has been around since the late 1950s, when computer researcher John McCarthy invented time sharing that allowed multiple users to access mainframes. This innovation would eventually result in the emergence of the Internet. After all, the Internet was still fairly archaic and not in wide-spread use.

Thus, with PCs, a company cloud network them together to enable collaboration and sharing of data and other resources. But as the Internet became more robust, there was a move to so-called cloud computing. One of the first business applications in this industry was developed by Salesforce.com, which made it possible for users to use the software through a browser. The downsides with cloud software are:

- With less control of the platform, there are more vulnerabilities to security and privacy lapses.
- Outages do happen and can be extremely disruptive and costly for enterprises that need reliability.
- Cloud computing is not necessarily cheap. In fact, one of the biggest complaints against Salesforce.com is the cost.

Besides the impact of Salesforce.com and other cloud applications companies, another critical development was Amazon.com's AWS platform. Launched in 2005, this allowed any company to build their own cloud-native applications. AWS essentially handles the complex administrative and infrastructure requirements like storage, security, compute, database access, content delivery, developer tools, deployment, IoT (Internet of Things), and analytics (there are currently more than 165 services). This means the development of applications can be much quicker.

The costs are generally lower and the fees are based on a per-use basis. With AWS, other mega tech firms were caught off guard and scrambled to develop their own cloud platforms. The two

most common ones include Microsoft's Azure and Google Cloud. In fact, many companies often use two or more of these in order to provide for redundancy (this is known as a multi-cloud strategy). The cloud also has different approaches, such as the following:

- **Public Cloud:** This is the model we've been covering so far in this chapter. That is, the cloud is accessed from remote servers, such as from AWS, Salesforce.com, and Microsoft. The servers have an architecture known as multitenant that allows the users to share a large IT infrastructure in a secure manner. This greatly helps to achieve economies of scale, which would not be possible if a company created its own cloud.
- **Private Cloud:** This is when a company owns the data center. True, there are not the benefits of the economies of scale from a public cloud. But this may not be a key consideration. Some companies might want a private cloud because of control and security.
- **Hybrid Cloud:** This is a blend of the public and private clouds. For example, the public cloud may handle less mission-critical functions.

As for RPA, the cloud has different implications and impacts. One is that a platform needs to deal with complex distributed applications, which can be difficult if a company develops custom programs on a cloud service. What's more, most RPA platforms actually started as on-premise software and generally did not transition to the cloud until recently. This does seem odd as cloud computing has been around for a while and appears to be the default approach for software companies. But developing cloud-native systems is not easy for RPA as there needs to be deep hooks across many applications and environments. In some cases, an on-premise RPA system may be loaded onto a cloud service like AWS. While there are benefits with this, it is not cloud native. This is because you will still need to upgrade and maintain the software.

WEB TECHNOLOGY

The mastermind of the development of the World Wide Web – which involved the use of hyperlinks to navigate web pages – was a British scientist, Tim Berners-Lee. He accomplished this in 1990. Although, it would not be until the mid-1990s, with the launch of the Netscape browser, that the Internet revolution was ignited. At the core of this was HTML or hypertext markup language, which was a set of commands and tags to display text, show colors, and present graphics. A key was that the system was fairly easy to learn and use, which helped to accelerate the number of web sites. For example, many of the commands in HTML involve surrounding content with tags, such as the following:

```
<strong>This is a Title</strong>
```

This means that the text is bold. Yet HTML would ultimately be too simple. So there emerged other systems to provide even richer experiences, such as with CSS (Cascading Style Sheets, which provides for borders, shadows, and animations) and JavaScript (this makes it possible to have sophisticated interactivity, say, with the use of forms or calculations). No doubt, RPA must deal with such systems to work effectively. This means it will have to take actions like identify the commands and tags so as to automate tasks.

PROGRAMMING LANGUAGES AND LOW CODE

A programming language allows you to instruct a computer to take actions. The commands generally use ordinary words like IF, Do, While, and Then. But there can still be lots of complexity, especially with languages that use advanced concepts like object-oriented programming. Some of the most popular languages today include Python, Java, C++, C#, and Ruby. To use an RPA system, you have to use some code – but it's not particularly difficult. It's actually known as low code. As the name implies, it is about using minimal manual input. For example, an RPA system has tools like drag-and-drop and visualizations to create a bot. This is not to imply that low code does not need some training. To do low code correctly, you will need to understand certain types of workflows and approaches.

OCR (Optical Character Recognition)

A key feature for an RPA platform is OCR (Optical Character Recognition), a technology that has actually been around for decades. It has two parts: a document scanner (which could even be something like your smartphone) and software that recognizes text. In other words, with OCR, you can scan an image, PDF, or even handwritten documents – and the text will be recognized. This makes it possible to manipulate the text, such as by transferring it onto a form or updating a database. There are definitely many challenges with effective OCR scanning, such as:

- The size of a font
- The shape of the text
- The skewness (is the text rotated or slanted?)
- Blurred or degraded text
- Background noise
- Understanding different languages

Because of all this, OCR in the early years was far from accurate. But over time, with the advances in AI algorithms, fuzzy logic, and more powerful hardware, the technology has seen great strides in accuracy rates, which can be close to 100%. The OCR can better capture the workflows by recognizing words and other visuals on the screen. So, even if there is a change of the location of these items, the RPA system can still identify them.

Something else: Automation involves large numbers of documents. Thus, OCR will greatly improve the processing. An example of this would be processing a loan. With OCR, a document will use OCR to extract information about a person's financial background, the information about the property, and any other financial details. After this, the RPA system will apply the workflows and tasks to process the loan, say, with applying various rules and sending documents to different departments and regulatory agencies.

Finally, even though RPA systems may have their own OCR, this may not necessarily be enough. Some industries and segments, such as healthcare, insurance, government and banking, still rely heavily on handwritten forms – all of which can be time-consuming and costly. But there are OCR systems that can help out, such as HyperScience. The software leverages sophisticated machine learning (ML) technology to quickly and accurately extract the information (understanding cursive writing, for example). But there are other capabilities like detecting fields on invoices and handling data reconciliation. Consider that HyperScience can automate up to 90% of the processing

DATABASES

At the heart of most applications is a database, which stores data that can be searched and updated. This is usually done by putting the information in tables (i.e., rows and columns of information). The dominant form is the relational database – developed in 1970 by IBM researcher E. F. Codd – that uses structured data. To interact with this, there is a scripting language called SQL (Structured Query Language), which was relatively easy to learn. It was not until the late 1970s that relational databases were commercialized, led by the pioneering efforts of Oracle. Then came a smattering of start-ups to seize the opportunity. But by the late 1980s, Oracle and SAP dominated the market for the enterprise. While relational databases proved to be quite effective, there were still some nagging issues. Perhaps the biggest was data sprawl. This describes when there is a growing number of tables that get proliferated across the organization. This often makes it extremely difficult to centralize the data, which can make it challenging to get a holistic view. Another problem was that relational databases were not cheap. And as new technologies came on the scene, such as cloud computing and real-time

mobile applications, it became more difficult to process the data. Given all this, there emerged various alternatives to relational databases. For example, there was the data warehouse that started as an open source project in the late 1990s from Doug Cutting. The technology would undergo various iterations, resulting in the development of Hadoop. Initially, Yahoo! used this to handle the Big Data demands from its massive digital platforms. Then other major companies, like Facebook and Twitter, adopted Hadoop. The key was that a data warehouse could make it possible to get a 360 degree view of data. The market has definitely seen lots of growth and change. Companies like Google, Amazon.com, and Microsoft have been investing heavily in data warehouse systems. There are also some fast-growing start-ups, like Snowflake, that are pushing the boundaries of innovation. In the meantime, there have been new approaches that have gone against the model for relational databases. They include offerings like MySQL (which is now owned by Oracle) and PostgreSQL. Yet these systems did not get enough traction in the enterprise. But there is one next-generation database technology that has done so: NoSQL. It also began as an open source project and saw tremendous growth. As of now, MongoDB has 14,200 customers across 100 countries and there have been over 70 million downloads. Where relational databases are highly structured, a NoSQL system is quite flexible. It's based on a document model that can handle huge amounts of data at petabyte scale. Another major secular trend is the transition of databases to the cloud. According to research from Gartner, about 75% will be migrated. For the most part, this should make it easier to allow for improved analytics and AI. And going forward, there is likely to be much innovation with database technology. Yet relational databases will remain the majority of what companies use – which also means that this will also be what RPA interacts with as well.

APIS (Application Programming Interface)

An API – which is the acronym for "application programming interface" – is software that connects two applications. For example, let's say you want to create a weather app. To get access to the data, you can setup an API, which often is fairly straightforward, such as by putting together a few lines of code to make data requests (say, for the city). By doing this, you will increase the speed of the development. APIs are pervasive in enterprise environments since they are so effective. They also have different structures. Although, the most common is a REST (representational state transfer) API. It's true that APIs can be used as a form of automation. Yet there are some things to keep in mind:

- The technology requires having people with technical backgrounds.

- The development of an API can take time and require complex integration. There is also the need for ongoing testing. However, there are third-party services that can help out.
- There must be a focus on maintaining an API (it's not uncommon for an API to break if there is a change in the structure).

Even if there is an off-the-shelf API available, there are still issues. One is metering, which means that you may be limited to a certain number of requests per day or hour. Or there may be higher pricing. Next, APIs can still have bugs and glitches, especially when in complex IT environments. Because of the difficulties, RPA has proven to be a very attractive alternative. Again, the development is much easier and there is less of a need for integration. But, interestingly enough, an RPA platform can be a vehicle for delivering advanced APIs within the enterprise

AI (ARTIFICIAL INTELLIGENCE)

A typical RPA system does not have much AI (Artificial Intelligence). The main reason is that there is a literal carrying out of tasks, which does not require any smart system. But as AI gets more powerful and accessible, RPA will increasingly start to use this powerful technology – which should greatly enhance the outcomes. AI is a software that ingests large amounts of data that is processed with sophisticated algorithms that help answer questions, detect patterns, or learn. Interestingly enough, AI is actually made up of a variety of subcategories (Figure 1.3 below shows a visual of this):

- **Machine Learning** : This is where a computer can learn and improve by processing data without having to be explicitly programmed. Machine learning is actually one of the oldest forms of AI and uses traditional statistical methods like k-nearest neighbor (k-NN) and the naive Bayes classifier.
- **Deep Learning** : While the roots of this go back to the 1960s, the technology was mostly an academic pursuit. It wasn't until about a decade ago that deep learning became a major force in AI. Some of the important factors for this included the enormous growth in data, the use of GPUs (graphics processing units) that provided for ultrafast parallel processing, and innovation in techniques like backpropagation. Deep learning is about using so-called neural networks – such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), and generative adversarial networks (GANs) – to find patterns that humans often cannot detect. NLP (natural language processing): This is AI that helps understand conversations. The most notable examples

of this include Siri, Cortana, and Alexa. But there are also many chatbots that focus on specific uses cases (say, with providing medical advice).

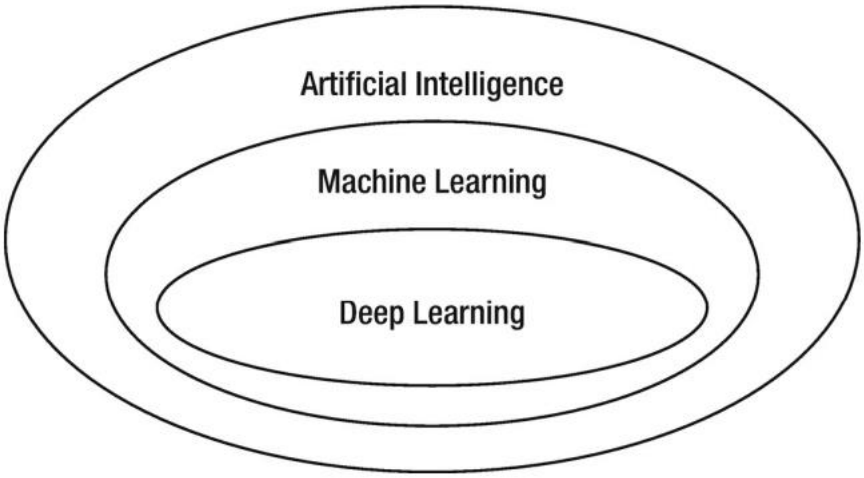


Fig.1.3: Subcategories of Artificial Intelligence

When it comes to AI, the excitement is often with the whiz-bang algorithms. But the reality is that this is often a small part of building a model. Keep in mind that it is absolutely essential to have high-quality data. And this usually means spending much time cleaning it up as well as weeding out outliers, which can be a tedious manual process. Another confusion about AI is, well, the impact from science fiction movies. These portrayals are about strong AI or AGI (artificial general intelligence), in which computers act like humans. Systems can think, engage in conversion, and walk around. But with today's technology, it's only about weak AI. This means that the applications are focused on narrow areas, such as helping predict weather or mining insights about customers. Besides, AI has some major issues, such as the following:

1. **Bias** : According to IBM: "Bad data can contain implicit racial, gender, or ideological biases. Many AI systems will continue to be trained using bad data, making this an ongoing problem." A real-world example of this is Amazon.com, which built a recruiting system for hiring programmers. The problem? It kept selecting males! Amazon.com did change the system– and the results did not change much. The inherent problem was that much of the data – which was based on incoming resumes – were from males. In other words, AI turned out to be the wrong approach. The good news is that Amazon.com recognized this and abandoned the project. But how many cases are there where companies don't? If so, they may be engaging in unwilling discrimination. This could mean the company will miss out on attracting good candidates. Even worse, there may be legal liability.
2. **Causation**: Humans have a strong grasp of this. We know what will happen if we use a hammer to hit a glass. It's pretty much instinctive. But AI is another matter. This

technology is really about finding correlations in data not causation – and this is a major limiting factor.

3. **Common Sense:** A human does not have to process many cases to understand certain rules of thumb. We just naturally understand them. But with AI, common sense has been extremely difficult to code because of the ambiguity and the lack of useful data for the seemingly infinite use cases.
4. **Black Box :** Deep learning can have an enormous number of layers and parameters. This means it can be nearly impossible for a person to understand why the model is generating certain results. True, this may not be a problem with facial recognition. But with applications in regulated industries, it could mean that deep learning is not viable. Consider that the deep learning systems are not allowed in financial services. Now there is more innovation in trying to find ways to understand deep learning outcomes – which is something called "*explainability*" – but the efforts are still in the nascent stages.
5. **Comprehension** : An AI system cannot truly understand what it is reading or observing. For example, if it read War and Peace, it would not be able to provide thoughts on the character development, themes, and so on.
6. **Static**: So far, deep learning has been mostly useful with constrained environments, such as with board games. There is a defined set of dimensions, objects, and rules – the kinds of things that computers work well with. It is also possible to conduct millions of simulations to learn. But of course, the real world is much more dynamic, open-ended, and chaotic.
7. **Conceptual Thinking**: AI cannot understand abstract ideas like justice, misery, or happiness. There is also a lack of imagination and creativity.
8. **Brain**: It's really a miracle of evolution. A typical brain has 86 billion neurons and trillions of synapses. And it only needs 50 watts a day to run! Modern computers can come nowhere matching this power. So if AI is to truly achieve real intelligence, there will need to be some dramatic breakthroughs.

Then so what about AI and RPA? It's certainly a major focus right now in the industry and we'll see many developments in the years ahead. But for now, it's important to keep some things in mind. First of all, there are two main types of data:

- Structured Data : This is data that is formatted (social security numbers, addresses, point of sale information, etc.) that can be stored in a relational database or spreadsheet.
- Unstructured Data : This is data that is unformatted (images, videos, voicemails, PDFs, emails, and audio files). For the most part, RPA uses structured data.

COGNITIVE AUTOMATION

It's often confused with AI – but the two concepts have different meanings. Consider cognitive automation to be an application of AI, actually. First of all, it is mostly focused on automation of the workplace or processes in business. Next, cognitive automation uses a combination of technologies like speech recognition and NLP. By doing this, the goal is to replicate human actions as best as possible, such as by analyzing patterns of workers and then finding patterns and correlations. Unlike other forms of AI, cognitive automation is usually effective with the use of much less data.

AGILE, SCRUM, KANBAN AND WATERFALL

Software development can be quite complex. Besides the technical aspects, there is a need to manage a team whose members may be located in different countries. In the meantime, technologies continue to evolve. What's often even harder is maintaining a software system as there is usually a need to add capabilities and upgrade the underlying technologies. Even some of the world's most talented software companies have faced monumental challenges. Just look at Microsoft. The company's Vista operating system took five years to develop. The reason for this was that Microsoft had too many silos within its organization and did not have sufficient collaboration. The result was that the development was glacially slow and there were persistent mishaps. The irony was that one team may have created an effective piece of code but it did not work with the complete system. In today's world, software development has become even more difficult because of the emergence of new platforms like the cloud and the hybrid cloud. This is why it's important to look at software management approaches.

- **Agile**: One is called Agile, which was created back in the 1990s (a big part of this was the publication of the Manifesto for Agile Software Development). The focus of this was to allow for incremental and iterative development, which begins with a detailed plan. This also requires much communication across the teams and should involve people from the business side of the organization.
- **Scrum:** This is actually a subset of Agile. But the iterations are done as quick sprints, which may last a week or two. This can help with the momentum of a project but also make a larger project more manageable
- **Kanban**: This comes from the Japanese word for visual sign or card (the roots of the system go back to Toyota's high-quality manufacturing processes). So yes, with Kanban, there is the use of visuals to help streamline the process.

- **Waterfall**: This is the traditional code development model, which goes back to the 1970s. The waterfall model is about following a structured plan that goes over each step in much detail. To help this along, there may be the use of a project management tool, say, a Gantt chart. While the waterfall approach has its advantages, it has generally fallen out of favor. Some of the reasons are as follows:
    - It can be tough to make changes
    - the process can be tedious
    - there is often a risk of a project being late.

DEVOPS

DevOps has emerged as a critical part of a company's digital transformation. The "Dev" part of the word is actually more than just about coding software. It also refers to the complete application process (such as with project management and quality assurance or QA). As for "Ops," it is another broad term, which encompasses system engineers and administrators as well as database administrators, network engineers, security experts, and operations staff. For the most part, DevOps has come about because of some major trends in IT. One is the use of agile development approaches (this was discussed earlier in the chapter). Next is the realization that organizations need to combine technical and operational staff when introducing new technologies and innovations. And finally, DevOps has proven effective in working with cloud computing environments. According to Atlassian, which is a leading developer of DevOps tools: "The bad news is that DevOps isn't magic, and transformations don't happen overnight. The good news is that you don't have to wait for upper management to roll out a large-scale initiative. By understanding the value of DevOps and making small, incremental changes, your team can embark on the DevOps journey right away."

FLOWCHARTS

Since an essential part of RPA is understanding workflows and systems, the use of flowcharts is common. It's usually at the core of the software application. With a flowchart, you can both sketch out the existing workflows of a department. And then from here, you can brainstorm ways of improving them. Then you can use the flowchart to design a bot for the automation. The flowchart is relatively simple to use and it also provides a quick visual way to understand what you are dealing with. As the old saying goes, a picture is worth a thousand words. So let's take a look at some of the basics:

- **Terminator**: This is a rectangle with rounded corners and is used to start and end the process, as seen in Figure below

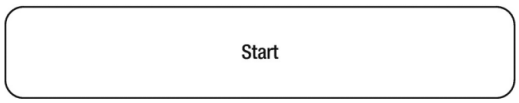


Fig. 1.4: This is a terminator, which starts and ends a flowchart.

- **Process**: This is represented by a rectangle. With this, there is only one next step in the process. Figure below shows an example:

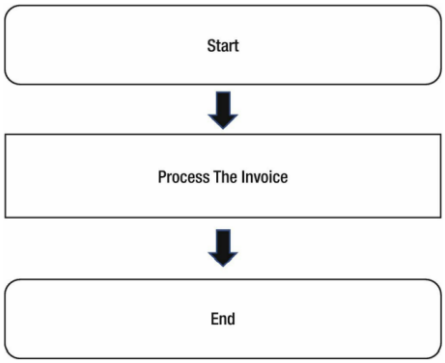
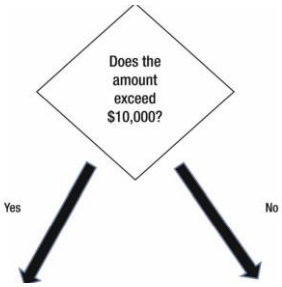


Fig. 1.4: This is a terminator, which starts and ends a flowchart.

- **Decision**: This is a square symbol that is at an angle. There will be at least two possible paths. Figure below is an example



----------**********----------

QUESTION BANK

Chapter 1: RPA Foundations:

1. What is RPA? List some of the tasks that can be performed by RPA bots.
2. List and Explain the Flavors of RPA.
3. Write a note on the History of RPA.
4. List and Explain the benefits of RPA.
5. List and Explain the downsides/drawbacks of RPA.
6. Write the note of the following:
    a. RPA compared with BPO
    b. RPA compared with BPM
    c. RPA compared with BPA

Chapter 2: RPA Skills:

1. Explain On-Premise Technology and the Cloud.
2. Explain Cloud Computing along with its drawbacks.
3. Explain the different Cloud Computing approaches
4. List the different RPA skills.
5. Explain Programming Language and Low Code.
6. Write a note on Databases.
7. Explain OCR feature used in RPA platforms.
8. Explain how API can be used in RPA.
9. Explain Artificial Intelligence.
10. Write a short note Machine Learning, Deep Learning and NLP.
11. List and explain the issues in Artificial Intelligence.
12. List and Explain the types of data.
13. Write a note on Cognitive Automation.
14. Explain Agile, Scrum, Kanban, and Waterfall.
15. Explain DevOps.
16. Explain Flowcharts.

----------**********----------